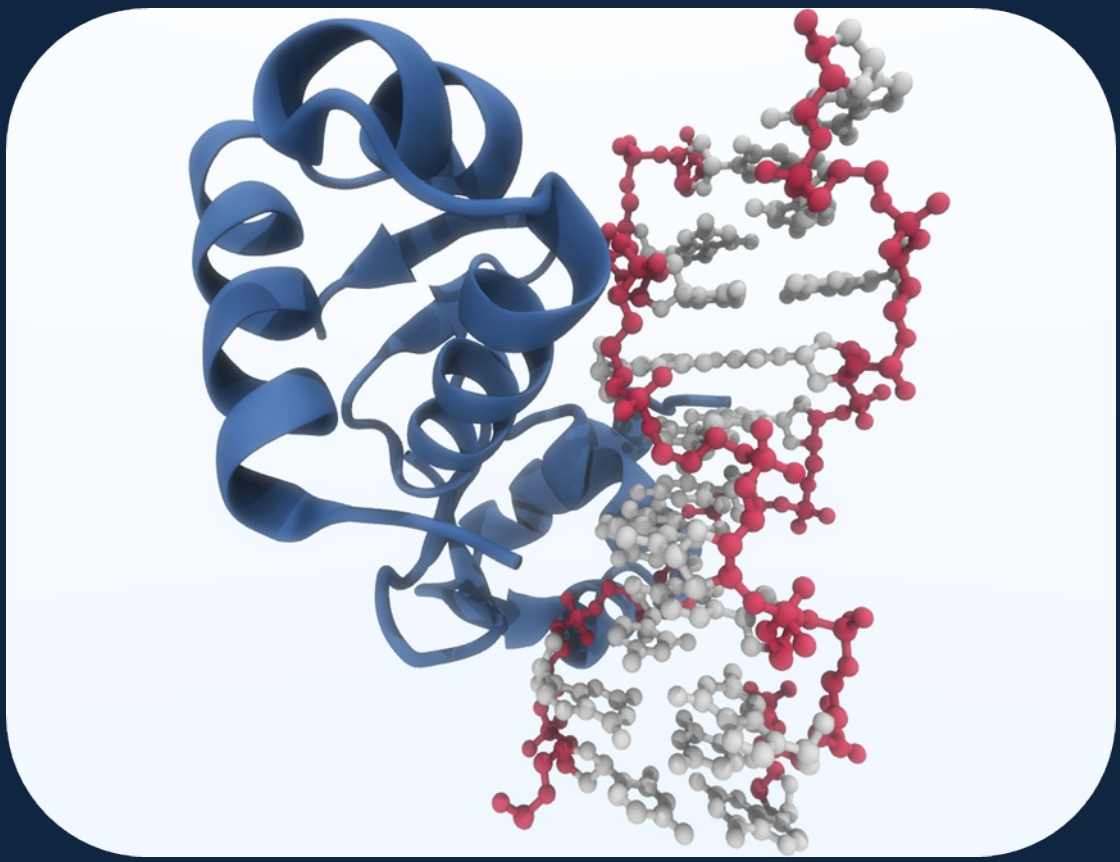
At the crossroad between innate immunity and RNA editing:

Structural and biochemical studies of the Zα domain family and the RNA editing enzyme ADAR1

Krzysztof Kuś



Dissertation presented to obtain the Ph.D. degree in Structural Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
July 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

At the crossroad between innate immunity and RNA editing:

Structural and biochemical studies of the Zα domain family and the RNA editing enzyme ADAR1

Krzysztof Kuś

Dissertation presented to obtain the Ph.D. degree in Structural Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Oeiras,
July 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER / UNL

Knowledge Creation

At the crossroad between innate immunity and RNA editing:

Structural and biochemical studies of the Zα domain family and the RNA editing enzyme ADAR1

Krzysztof Kuś

Supervisor: Alekos Athanasiadis

Cover image: Zα ORF112 dimer bound to Z-DNA

Declaration/Declaração

I declare that this dissertation is a result of my own research and carried out between April 2011 and October 2015 in the laboratory of Dr. Alekos Athanasiadis, Instituto Gulbenkian de Ciência in Oeiras, Portugal. Results presented in Chapter 2 has been published in the Journal of Virology entitled "Crystal structure of a poxvirus-like zalpha domain from cyprinid herpesvirus 3", Ana R. Tomé[§], Krzysztof Kuś[§], Silvia Correia, Lara M. Paulo, Sónia Zacarias, Matteo de Rosa, Delio Figueiredo, R. Michael E. Parkhouse and Alekos Athanasiadis ([§]equally contributed to this work). Chapter 3 was included in the publication: "The Structure of the *Cyprinid herpesvirus 3* ORF112-Zα·Z-DNA Complex Reveals a Mechanism of Nucleic Acids Recognition Conserved with E3L, a Poxvirus Inhibitor of Interferon Response" in the Journal of Biological Chemistry authored by Krzysztof Kuś, Krzysztof Rakus, Maxime Boutier, Theokliti Tsigkri, Luisa Gabriel, Alain Vanderplasschen and Alekos Athanasiadis. Chapters 4 and 5 contain unpublished data.

Declaro que esta dissertação é o resultado do meu próprio trabalho desenvolvido entre Abril de 2011 e Outubro de 2015 no laboratório do Dr. Alekos Athanasiadis, Instituto Gulbenkian de Ciência em Oeiras, Portugal. Os resultados do Capítulo 2 foram publicados no Journal of Virology como "Crystal structure of a poxvirus-like zalpha domain from cyprinid herpesvirus 3", Ana R. Tomé[§], Krzysztof Kuś[§], Silvia Correia, Lara M. Paulo, Sónia Zacarias, Matteo de Rosa, Delio Figueiredo, R. Michael E. Parkhouse and Alekos Athanasiadis ([§]igual contribuição). O capítulo 3 foi integrado num artigo publicado como "The Structure of the *Cyprinid herpesvirus 3* ORF112-Zα·Z-DNA Complex Reveals a Mechanism of Nucleic Acids Recognition Conserved with E3L, a Poxvirus Inhibitor of Interferon Response" no Journal of Biological Chemistry com autoria de Krzysztof Kuś, Krzysztof Rakus, Maxime Boutier, Theokliti Tsigkri, Luisa Gabriel, Alain Vanderplasschen and Alekos Athanasiadis. Os capítulos 4 e 5 contêm dados que não foram publicados.

Financial Support/ Apoio Financeiro

This dissertation had the financial support from FCT, doctoral fellowship #SFRH/BD/51626/2011 and Fundação Calouste Gulbenkian.

Esta dissertação teve o apoio financeiro da FCT, bolsa de doutoramento #SFRH/BD/51626/2011 e da Fundação Calouste Gulbenkian.

Acknowledgments

This amazing chapter of my life is approaching its end. There are so many people that helped me over past years.

First of all, I would like to express my deepest gratitude to my supervisor, Alekos Athanasiadis, not only for accepting me in his laboratory but also providing with an immense amount of support. It was an honor and pleasure to be a member of his laboratory. He introduced me to the world of crystallography and taught me how to approach science from different angles. Moreover, I am extremely grateful for all his patience and all fruitful discussions.

I was lucky to have such great lab mates (former and present): Ana Rita, Daniela, Diogo, Filipe, Klito, Lara, Luisa, Margarida, Matteo. I am also really glad that I could share part of my work with Ana Rita. It was fun to develop one of the projects collaborating with Luisa. Thank you all for great moments we had.

I would like to thank all people that are/were responsible for IGC Ph.D. Programme in Integrative Biomedical Sciences, especially: Thiago Lopes Carvalho, Élio Sucena, Manuela Cordeiro. I owe a debt of gratitude to my Thesis Committee: Mónica Bettencourt-Dias and Jorge Carneiro.

Nothing could be possible without the support of my family.

Finally, I am extremely grateful to Ewa, not only for her presence and constant support but also for all discussions and suggestions over this thesis.

Summary

Detection of foreign nucleic acids is a crucial component of innate immunity against viruses. For that purpose, cells are equipped with a repertoire of cytoplasmic and membrane-bound DNA/RNA sensors. Recognition of non-self nucleic acids leads to the activation of downstream pathways dedicated to elimination or containment of infection. One of the strategies to counteract viral infections is the production of interferons, the cytokines which alert neighboring cells about potential threats. Among the many sensors and effectors engaged in interferon responses, there is a small family of proteins with Z-DNA/Z-RNA binding domains. These domains, called Zα domains, can interact and stabilize the dynamically formed left-handed conformation of both DNA (Z-DNA) and RNA (Z-RNA). The exact role of Zα domains awaits to be elucidated but a growing body of evidence supports their involvement in the detection of foreign nucleic acids in the cell cytoplasm. On the other hand, pathogens seem to deploy Zα domains to avoid recognition by an innate immune system of vertebrates. Until recently, four proteins with Zα domains were identified: PKZ, DAI, ADAR1 (encoded in the vertebrate genomes) and the pox viral E3L.

PKZ, expressed in some fish species, is an orthologue of the well-studied Protein Kinase RNA-activated (PKR). PKR is one of the important elements of the innate immunity directed towards the recognition of the double-stranded RNAs (dsRNA) in vertebrate species. PKR is a modular protein which consists of two dsRNA-binding domains and a kinase domain. After detection of dsRNA, PKR dimerizes and autophosphorylates which leads to

the activation of the kinase. The activated PKR phosphorylates the eukaryotic initiation factor 2α (eIF2a) which results in the inhibition of protein translation. PKZ differs from PKR by harboring two Zα domains instead of dsRNA binding domains. PKZ, similarly to PKR, can shut down protein translation but most likely it recognizes a different subset of nucleic acids. Inhibition of protein translation results in the accumulation of stalled, pre-initiation complexes to cytoplasmic aggregates – the stress granules. These structures are postulated to serve as a sorting machinery directing mRNAs for storage, degradation or re-initiation of protein translation. Zα domains (from ADAR1, DAI, E3L) were shown to localize to stress granules and this process requires Z-DNA/Z-RNA binding activity of these domains.

ORF112 from cyprinid herpesvirus 3 was a protein without assigned function. Using homology-based searches, we found that ORF112 has a Zα domain. Intriguingly, ORF112 is encoded in the genomes of viruses which infect fish expressing PKZ. Therefore, ORF112 can be a putative inhibitor of PKZ and may sequester viral nucleic acids and prevent their detection. In this work, we demonstrated that Zα ORF112 interacts with Z-DNA *in vitro*. We then solved the crystal structure of the free protein which confirmed that Zα ORF112 has the typical Zα fold. This structure revealed that Zα ORF112 forms a homodimer through a domain-swapping mechanism: the C-terminal parts of the monomers are engaged in a reciprocal exchange of ten amino acids. The formation of such a dimer was not detected before for other members of the Zα domain family. Follow-up work showed that dimerization is induced by the presence of sulfate ions which mimic the phosphates of DNA backbone.

Finally, we determined the crystal structure of Zα ORF112 in complex with a T(CG)9 oligonucleotide at 1.5 Å. The structural data

confirmed that the bound DNA is in the left-handed conformation. We discovered that Zα ORF112 forms an on-DNA dimer which confers additional stabilization of the Z-conformation, a finding that is supported by our data from isothermal titration calorimetry. Moreover, we provide experimental evidence that ORF112 is targeted to stress granules similarly to other proteins with Zα domains. These results affirm that Zα ORF112 shares structural and functional properties with the other Z-DNA/Z-RNA binding domains. We predict that Zα ORF112 can be important for the pathogenesis of cyprinid herpesviruses (as previously shown for Zα E3L from poxviruses). Cyprinid herpesvirus 3 is a pathogen which is primarily responsible for massive mortality in the wild and agriculture populations of koi and common carp. The provided detailed information about the recognition of the nucleic acids by Zα ORF112 can serve as a basis for engineering of attenuated viruses and vaccine development.

As mentioned above, Zα domains localize to the stress granules and the mutations of residues involved in Z-DNA/Z-RNA binding lead to the loss of accumulation of Zα domains in these transient cytoplasmic structures. These findings suggest that Zα domains localize to stress granules due to their ability to interact with RNA present in these structures. As the intracellular substrates of the Zα domains are largely unknown, we decided to devise a methodology to capture and sequence the host nucleic acids bound to Zα domains. As our model, we selected the DNA-dependent activator of IFN-regulatory factors (DAI) with two functional Zα domains. We prepared cell lines stably expressing Zα DAI domains as a fusion protein with GFP (and controls with mutated Zα not able to interact with Z-DNA/Z-RNA). We verified that wild-type, but not mutant Zα DAI, domains are enriched in stress granules. Subsequently, we developed a modified version of the

cross-linking and immunoprecipitation (CLIP) protocol to prepare and sequence wild-type and mutant Zα-bound nucleic acids. Despite the promising preliminary results, further improvements to the protocol are required. Once fully optimized, this method should reveal the identity of the intracellular nucleic acid targets of these domains. This will further aid our understanding of biological processes involving Zα domains.

The prototypic Zα domain was found in the RNA editing enzyme ADAR1 (double-stranded RNA-specific adenosine deaminase). This protein is modular and, apart from Zα domains, it consists of double-stranded RNA-binding domains (dsRBDs) and a catalytic domain. ADAR1 belongs to the bigger family of ADARs – enzymes responsible for deamination of adenosines to inosines (A-to-I editing) in the context of the duplex RNA. A-to-I editing can affect many biological processes (recoding of amino acids, splicing and RNA interference) because inosines are recognized as guanosines by the cellular machinery. All ADARs can be either promiscuous (edit up to 50% of adenosines in perfect duplexes) or site-specific (target only specific adenines in the imperfect structures). The site-selective editing is modulated by the presence of features like bulges, loops, mismatches in the dsRNA, but the exact mechanisms by which ADARs recognize dsRNA structures are not well understood. Thus, we aimed at structural characterization of the RNA recognition by the catalytic domain of ADAR1. In this thesis, we describe attempts to create an expression system for the catalytic domain of ADAR1. Additionally, as dsRBDs play an important role in the substrate recognition by ADAR1 we have performed biochemical characterization of the first and the third dsRBDs of human ADAR1. Our structural and biochemical approaches were extended by computational studies of potential ADAR substrates. This resulted in the development of an *in silico*

method to search for co-occurrence of silent point mutations and RNA editing sites in the predicted RNA secondary structures. Our aim was to provide data that can be used to test how silent point mutations modulate the editing levels.

Altogether, our work extended a catalog of proteins with Zα domains by the new member – ORF112. We provided important insights into how the newly discovered Zα ORF112 domain interacts with nucleic acids. Our data suggest that ORF112 serves as an inhibitor of fish immune responses by protecting viral nucleic acids from being detected by host PKZ. Moreover, we performed initial biochemical and computational characterization of ADAR1 domains and their potential substrates. Finally, we developed an *in vitro* protocol to capture nucleic acids bound to Zα domains, which in the future may reveal novel functions of Z-DNA/Z-RNA binding domains.

Sumário

A deteção de ácidos nucleicos exógenos é um componente essencial da imunidade inata contra vírus. Para esse efeito, as células possuem no seu repertório sensores citoplasmáticos e membranares de ADN/ARN. O reconhecimento dos ácidos nucleicos exógenos traduz-se na ativação das vias de sinalização a jusante, de forma a eliminar ou a conter a infeção. A produção de interferões é uma das estratégias para neutralizar a infeção viral. Os interferões são citocinas que alertam as células vizinhas para potenciais ameaças. Entre os diversos sensores e efetores envolvidos nos mecanismos de defesa mediados por interferões, existe uma família de proteínas com domínios que se ligam a Z-ADN/Z-ARN. Estes domínios, designados domínios Zα, podem interagir e estabilizar a conformação em hélice virada para a esquerda que se forma dinamicamente a partir do ADN (Z-ADN) e do ARN (Z-ARN). O papel exato dos domínios Zα ainda não foi elucidado, mas cada vez mais provas suportam o seu envolvimento na deteção de ácidos nucleicos exógenos no citoplasma da célula. Por outro lado, os agentes patogénicos parecem utilizar domínios Zα para evitar o reconhecimento pelo sistema imune inato de vertebrados. Até recentemente, quatro proteínas foram identificadas com domínios Zα: PKZ, DAI, ADAR1 (codificadas nos genomas de vertebrados) e E3L (presente em vírus da família *Poxviridae*).

PKZ, expressa em algumas espécies de peixes, é uma proteína ortóloga à proteína cinase activada por ARN (PKR). A PKR é um dos elementos importantes da imunidade inata no que diz

respeito ao reconhecimento de ARN em cadeia dupla (dsRNA) em vertebrados. A PKR é uma proteína modular composta por dois domínios de ligação a dsRNA (dsRBD) e o domínio cinase. Após a deteção de dsRNA, PKR dimeriza e autofosforila-se, o que resulta na ativação da cinase. Por sua vez, a PKR ativada fosforila o fator de iniciação eucariótico 2α (eIF2a), que conduz à inibição da síntese proteica. PKZ difere de PKR por albergar dois domínios Zα em vez de domínios de ligação a dsRNA. Tal como PKR, PKZ pode terminar a tradução da proteína mas provavelmente através do reconhecimento de um subconjunto diferente de ácidos nucleicos. A inibição da tradução da proteína resulta na acumulação de complexos de pré-iniciação de tradução de proteínas, estagnados em agregados citoplasmáticos – os grânulos de estresse. Postula-se que estas estruturas funcionam como uma máquina de triagem que dirige o ARN mensageiro (mRNA) para armazenamento, degradação ou reiniciação da tradução da proteína. Demonstrou-se que domínios Zα (de ADAR1, DAI, E3L) localizam-se em grânulos de estresse, devido à sua atividade de ligação a Z-ADN/Z-ARN.

A proteína ORF112, identificada no herpesvirus 3 ciprinídeo, não tinha nenhuma função atribuída. Através de pesquisas baseadas em homologia, descobrimos que a ORF112 tem um domínio Zα. Curiosamente, a ORF112 está codificada nos genomas de vírus que infetam peixes expressando PKZ. Portanto, a ORF112 pode ser um potencial inibidor da PKZ, sequestrando os ácidos nucleicos virais e impedindo o seu reconhecimento. Neste trabalho demonstramos que Zα ORF112 interage com Z-ADN *in vitro*. Em seguida, resolvemos a estrutura da proteína livre, que confirmou que Zα ORF112 tem a conformação típica dos domínios Zα. Esta estrutura revelou que Zα ORF112 forma um homodímero através de um mecanismo de troca de domínio: as partes C-terminais dos monómeros estão envolvidas numa troca recíproca de dez

aminoácidos. A formação de tal dímero nunca tinha sido detetada em nenhum outro membro da família de proteínas que possuem domínios Zα. Mais tarde demonstrou-se que a dimerização é induzida pela presença de iões sulfato, que imitam os fosfatos do esqueleto de ADN.

Finalmente, determinámos a estrutura do domínio Zα em complexo com o oligonucleótido $T(CG)_9$ a 1,5 Å. Os dados estruturais confirmaram que o ADN complexado se apresenta na conformação em hélice virada para a esquerda. Descobrimos que Zα ORF112 dimeriza quando ligado ao ADN, o que confere estabilidade adicional à conformação de Z-ADN, uma conclusão que é apoiada pelos nossos dados de calorimetria de titulação isotérmica. Além disso, evidenciamos experimentalmente que ORF112 é direcionado para grânulos de estresse de forma semelhante a outras proteínas com domínios Zα. Estes resultados permitem afirmar que Zα ORF112 compartilha as características estruturais e funcionais com os outros domínios que se ligam a Z-ADN/Z-ARN. Prevemos que Zα ORF112 pode ter um papel importante na patogénese do herpesvirus ciprinídeo (tal como anteriormente demonstrado para o domínio Zα da proteína E3L presente em vírus da família *Poxviridae*). Herpesvirus 3 ciprinídeo é um agente patogénico responsável principalmente pela mortalidade em populações selvagem e em cativeiro de carpas, incluindo a carpa comum e a carpa Koi. A informação detalhada fornecida sobre o reconhecimento dos ácidos nucleicos por Zα ORF112 pode servir de base para engenharia de vírus atenuados e o desenvolvimento de vacinas.

Como foi mencionado acima, os domínios Zα localizam-se em grânulos de estresse e as mutações de resíduos de aminoácidos envolvidos em ligação a Z-ADN/Z-ARN resultam na perda de acumulação de domínios Zα nestas estruturas transientes

no citoplasma. Estes resultados sugerem que os domínios Zα localizam-se em grânulos de estresse, devido à sua capacidade de interagir com o ARN presente nestas estruturas. Como os substratos intracelulares dos domínios Zα são em grande parte desconhecidos, decidiu-se criar uma metodologia para capturar e sequenciar os ácidos nucleicos da célula hospedeira ligados aos domínios Zα. Como modelo, selecionámos a proteína ADN-dependente activador de factores de IFN-regulação (DAI), que possui dois domínios Zα funcionais. Preparámos linhas celulares que expressam de forma estável os domínios Zα de DAI como uma proteína de fusão com GFP (assim como controlos com os domínios Zα mutantes, incapazes de interagir com Z-ADN/Z-ARN). Observámos que a versão selvagem, mas não a mutante, dos domínios Zα DAI se localiza preferencialmente em grânulos de estresse. Posteriormente, desenvolvemos uma versão modificada do protocolo de cross-linking e imunoprecipitação (CLIP) para preparar e sequenciar os ácidos nucleicos ligados aos domínios Zα, quer a versão selvagem, quer a versão mutante. Apesar dos resultados preliminares serem promissores, o protocolo necessita ainda de algumas otimizações adicionais. Devidamente otimizado, este método deverá desvendar a identidade dos ligandos alvo de ácidos nucleicos intracelulares para estes domínios. Isso irá ajudar a compreensão dos processos biológicos que envolvem domínios Zα.

O primeiro domínio Zα a ser identificado foi encontrado na enzima ADAR1 (adenosina deaminase específica de ARN em cadeia dupla). Esta proteina é modular e, além dos domínios de ligação de Z-ADN/Z-ARN, contém domínios de ligação a ARN de cadeia dupla (dsRBD) e um domínio catalítico. A ADAR1 pertence à família de ADARs - enzimas responsáveis pela desaminação da adenosinas para inosinas (edição de A-para-I) em ARN de cadeia dupla. Edição de A-para-I pode afetar muitos processos biológicos

(recodificação de aminoácidos, *splicing*, ARN interferente) porque as inosinas são reconhecidas como guanosinas pela maquinaria celular. Todas as ADARs podem atuar de forma promíscua (editando até 50% das adenosinas presentes em duplexes perfeitos) ou específica (editando só adenosinas específicas em estruturas imperfeitas). A edição seletiva é modulada pela presença de bulges, loops e desemparelhamentos do ARN de cadeia dupla, mas os mecanismos exactos pelos quais as ADARs reconhecem as estruturas de dsRNA não são bem compreendidos. Por isso, resolvemos caracterizar os detalhes estruturais envolvidos no reconhecimento de ARN pelo domínio catalítico da ADAR1. Nesta tese, descrevemos tentativas de criar um sistema de expressão para o domínio catalítico da ADAR1. Além disso, como os dsRBDs desempenham um papel importante no reconhecimento do substrato pela ADAR1, caracterizámos bioquimicamente o primeiro e o terceiro dsRBDs da ADAR1 humana. As nossas abordagens estruturais e bioquímicas foram complementadas pelos estudos computacionais de identificação de potenciais substratos das ADARs. Isto resultou no desenvolvimento de um método *in silico* para procurar a coocorrência de mutações pontuais silenciosas e locais de edição de ARN nas estruturas secundárias de ARN previstas. O nosso objetivo era fornecer dados que pudessem ser usados para testar como mutações pontuais silenciosas modulam os níveis de edição.

Ao todo, o nosso trabalho expandiu o repertório de proteínas com domínios Zα com a adição de um novo membro - ORF112. Fornecemos informações importantes sobre a forma como o recém-descoberto domínio Zα ORF112 interage com ácidos nucleicos. Os nossos dados sugerem que a ORF112 funciona como um inibidor das respostas imunes de peixe, protegendo os ácidos nucleicos virais de serem detetados pela proteína PKZ do hospedeiro. Além disso, realizámos a caracterização bioquímica e computacional

inicial de domínios ADAR1 e dos seus substratos potenciais. Finalmente, desenvolvemos um protocolo *in vitro* para identificar ácidos nucleicos ligados a domínios *Zα* que, no futuro, poderá revelar novas funções de domínios que se ligam a Z-ADN/Z-ARN.

Contents

List of Figures

List of Tables

Abbreviations

ADAR	double-stranded RNA-specific adenosine deaminase
CLIP	cross-linking and immunoprecipitation
CyHV3	cyprinid herpesvirus 3
DAI	DNA-dependent activator of IFN-regulatory factors
dsRBD	double-stranded RNA binding domain
dsRNA	double-stranded RNA
IFN	interferon
ORF112	open reading frame 112
PKR	Protein Kinase RNA-activated
PKZ	PKR-like protein kinase containing a Z-DNA binding domain
SGs	stress granules
Zα	Z-DNA/Z-RNA binding domain
Z-DNA	left-handed DNA
Z-RNA	left-handed RNA

1

General Introduction

1.1 Detection of nucleic acids in the innate immunity

All living organisms interact with each other, and the line between ally or enemy, relative or stranger can be really thin. As discrimination between self and non-self is critical for the survival and proliferation of organisms, many mechanisms have evolved to ensure such a distinction. Important components of the innate immunity are the receptors for conserved foreign molecules i.e. nucleic acids. The nucleic acid recognition of self and non-self can be based on the modifications, localization, sequence or even structure. Despite the fact that sensing of nucleic acids expands the number of detected microbes, the imbalances in the recognition of these universal information carriers can have adverse effects on the host causing autoimmunity disorders. Clearly, our (*Homo sapiens*) interests are biased towards understanding how our (or, at least, eukaryotic) cells detect DNA/RNA of viral, bacterial and fungal origin.

1.1.1 The vertebrate innate immunity

The immunity of vertebrates can be divided into two branches: adaptive and innate immunity, which are complementary and highly connected. The adaptive immunity (also known as acquired) involves a network of cells producing antibodies and creating immunological memory, which would target the pathogen with great specificity during the infection or/and on the next encounter. Therefore, the adaptive immunity requires time to respond to the invasion of the bacteria, virus and fungi. As the first line of defense, organisms (not only vertebrates) have in their disposition a certain number of mechanisms which are present at the time of pathogen attack. The innate immunity system includes physical

barriers (like skin or cell wall of plants), phagocytic cells (e.g. macrophages) and the complement system and secreted molecules (e.g. lysozyme) (Elias, 2007). The innate immunity system is also represented by a number of receptors named pattern-recognition receptors (PRRs). PRRs have the ability to recognize molecular signatures of pathogens - named pathogen-associated molecular patterns (PAMPs). Many of PAMPs are only found in microbes (like LPS /lipopolysaccharide/, peptidoglycan, flagellin, lipoteichoic acid) and help to distinguish between self and non-self. On the contrary, the detection of nucleic acids is more problematic as these molecules are the universal information carriers. Undoubtedly, the recognition of nucleic acids largely increases the number of microbes detected by the host. Generally, upon binding to nucleic acids, PRRs would trigger downstream pathways activating many different responses (i.e. cytokine and interferon production, translation shut-down). Nucleic acids PRRs can be classified according to the cellular localization: membrane-bound and cytoplasmic (Wu and Chen, 2014).

1.1.2 Membrane-bound nucleic acids sensors

The group associated with membranes (endosomal and lysosomal compartments) is represented by Toll-like receptors (TLRs). TLRs belong to the family of transmembrane receptors and consist of the extracellular domain with many leucine-rich repeats (LRRs), the transmembrane domain, and the cytosolic signaling domain (named Toll/IL-1 receptor homology TIR) (Kawai and Akira, 2010). Currently, four TLR representatives (in different species) are known to participate in the nucleic acid detection: TLR3, TLR7, TLR8 and TLR9. TLR3 has been shown to respond to the dsRNA with the requirement of minimum 45 bp for the signal

transduction (Alexopoulou *et al.*, 2001; Botos *et al.*, 2009). Both TLR7 and TLR8 have been demonstrated to be a species specific part of ssRNA detection system (Lund *et al.*, 2004; Heil *et al.*, 2004). TLR7 performs its functions in mice whereas TLR8 is found in humans. Finally, TLR9 has been implicated in the recognition of CpG DNA (Hemmi *et al.*, 2000). Almost all TLRs (except TLR3) upon binding to cognate substrates recruit (directly or through other adaptors) MyD88 (Myeloid differentiation primary response gene 88). The TLR-MyD88 complex activates NF-κB or MAPK (the mitogen-activated protein kinase) pathway. These pathways trigger translocation of transcription factors to the nucleus which drive transcription of pro-inflammatory cytokines and interferons (De Nardo, 2015).

1.1.3 Cytoplasmic nucleic acids sensors

TLRs scan the endosome and lysosome milieu and trigger cellular responses to nucleic acids of phagocytized microbes and their constituents. Nevertheless, a vast number of viruses and some bacteria (i.e. Chlamydia) replicate in the cytoplasm. To assure that the cytoplasm is also protected cells have evolved an arsenal of soluble nucleic acid sensors that upon binding to cytosolic nucleic acids trigger down-stream pathways. Cytosolic PRRs can be grouped into two categories depending on the nature of nucleic acids recognized (RNA or DNA). The following paragraphs are intended to give a brief overview of the cytoplasmic RNA or DNA sensors.

Detection of DNA

Eukaryotic cells are compartmentalized with membranous organelles. The nucleus is one of the obvious features setting

the distinction between eukaryotes and prokaryotes. The nucleus contains the genetic material of the cell and is encapsulated by the nuclear envelope which separates DNA from the cytoplasm. Therefore, it is generally accepted that under normal conditions the cytoplasm of the eukaryotic cell is depleted from DNA. As a result, DNA found in the cytoplasmic and endosomal compartment would be interpreted as a foreign nucleic acid. Endosomal TLR9 can be activated by CpG DNA but it is primarily expressed in the immune-related cells. However, it has been demonstrated that other cell types can produce interferons when exogenous DNA is delivered to their cytoplasm (Stetson and Medzhitov, 2006; Ishii *et al.*, 2006). This was an indication that a TLR9-independent pathway exists and that the recognition of DNA may occur in the cytoplasmic compartment.

The first documented cytosolic DNA sensor - DAI (DNA-dependent activator of IFN-regulatory factors, known also as ZBP1/DLM-1) was postulated to respond to B-DNA poly(dA-dT)·poly(dT-dA)) and to a lesser extent to poly(dG-dC)·poly(dC-dG) (which may adopt a left-handed helical conformation, Z-DNA). DAI has two Z-DNA/Z-RNA binding domains (Zα domains) and C-terminal region termed D3. More detailed description of this protein is provided in the **Paragraph 1.3.2**.

Another consequence of the DNA detection by PRRs is the production of signaling molecules. Sensing of dsDNA by cyclic GMP-AMP synthase (cGAS) triggers a conformational change which leads to the induction of the enzymatic activity of the protein (Civril *et al.*, 2013). As a result, cGAS synthesizes cGAMP (cyclic guanosine monophosphate–adenosine monophosphate) which is a secondary messenger that binds to STING receptor and initiates a cascade leading to the production of IFNβ (Wu *et al.*, 2013). cGAMP can also be transported to the neighboring cells via gap

junctions and serve as an alert molecule. Moreover, STING can directly respond to the secondary messengers produced by the large number of bacterial species (Burdette *et al.*, 2011). Detection of the cytoplasmic dsDNA not only induces the transcription of type I IFNs or inflammatory cytokines but also the processing of the interleukins. AIM2 (absent in melanoma 2), a member of PYHIN family, is responsible for the assembly of the inflammasome. AIM2, bound to dsDNA of various sources (i.e. plasmids, poly(dA-dT), vaccinia virus DNA), oligomerizes and gains the ability to interact with ASC (Apoptosis-associated speck-like protein containing a CARD). ASC, complexed with AIM2, recruits procaspase-1 via its CARD domains. These events trigger the autoactivation of caspase-1, which is responsible for the maturation of IL-1β and IL-18. Interestingly, another member of PYHIN family, IFI16, was reported to either act through the inflammasome pathway (in response to DNA in the nucleus) (Kerur *et al.*, 2011) or to activate the STING-mediated production of IFNs (Unterholzner *et al.*, 2010).

Recognition of RNA

Double-stranded RNA (dsRNA) is produced during replication of many viruses (either as a genome or as byproducts of genome replication). The presence of dsRNA in the cytoplasm is a danger signal (a sign of viral infection) leading to protein translation arrest. Protein kinase R (PKR) is one of the best studied cytoplasmic sensors of dsRNAs. PKR not only can detect nucleic acids but also acts as an effector protein. When PKR binds dsRNA by its two dsRNA-binding domains, it forms a dimer and undergoes autophosphorylation leading to its activated form. Subsequently, PKR phosphorylates eukaryotic initiation factor 2α (eIF2α) (Taghavi and Samuel, 2013). The modification of eIF2α is responsible for a global protein translation shut-down. The

accumulation of the resulting stalled ribosomes leads to the formation of the stress granules (SGs) - dynamic protein-RNA assemblies (see **Figure 1.1**). The next paragraph provides a more detailed description of the potential involvement of SGs in the innate immunity with emphasis on the viruses (**Section 1.1.4**).

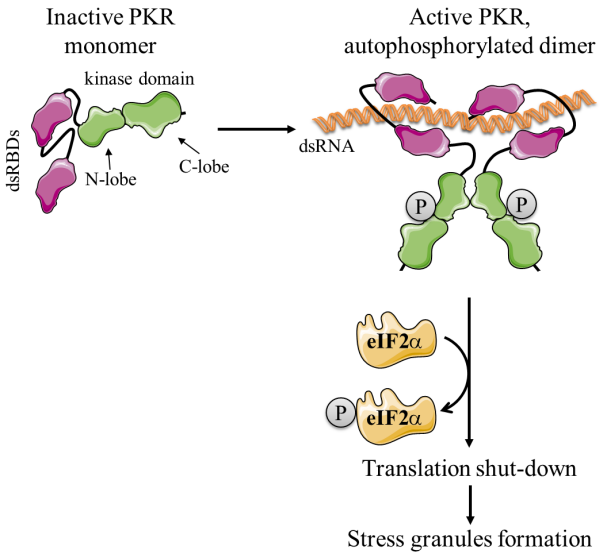


Figure 1.1: Detection of dsRNA by PKR leads to translation shut-down. PKR upon binding to cytoplasmic dsRNA (via dsRBD domains), autophosphorylates and dimerizes. Consequently, active PKR phosphorylates eIF2α a modification responsible for the arrest of protein translation. The resulting stalled pre-initiation ribosome complexes aggregate into cytoplasmic stress granules.

Another class of molecules able to recognize RNA in the cytoplasm are proteins belonging to the family of RIG-I-like receptors (RLRs). RLRs are members of DExD/H box ATP RNA helicases and currently three members are known to be involved in innate dsRNA sensing: RIG-I (retinoic acid-inducible gene 1), MDA5 (melanoma differentiation-associated protein 5) and LGP2 (laboratory of genetics and physiology 2). RIG-I responds with the highest affinity to RNA with 5'-triphosphate with blunt-ended, 5' base-paired region (where dsRNA can be intra- or intermolecular). It should be noted that a 5'-triphosphate is an important mark of the foreign nucleic acids in the cytoplasm as the

host nucleic acids are processed to be devoid of such modification or possess additional marks (Goubau *et al.*, 2013). In contrast, MDA5 seems to bind long dsRNA which is a hallmark of the infection with positive-strand RNA viruses. When MDA5 or RIG-I detect ligand RNA, they associate with the mitochondrial adaptor protein MAVS through a CARD motif (caspase activation and recruitment domain). As a consequence, the downstream pathways mediated by NF-κB or IRF3 are triggered and interferons are produced (Habjan and Pichlmair, 2015). Interestingly, some PRRs can amplify interferon responses by producing intermediates which activate other cytoplasmic sensors. One such example is OAS (oligoadenylate synthase), which after binding dsRNA produces 2'-5' linked oligoadenylates which in turn activate RNase L (Chakrabarti *et al.*, 2011). Then RNase L cleaves mRNA molecules of the host and viral origin rendering them substrates for RLRs. In some instances, the detection of DNA converges with the RNA sensing pathways. It has been demonstrated that RNA polymerase III (PolIII) can transcribe DNA rich in AT repeats into a 5'triphosphate dsRNA - a substrate for RIG-I (Chiu *et al.*, 2009).

1.1.4 Stress granule formation and antiviral innate immunity

The assembly of RNA-protein granules is a part of the normal cellular functions (i.e. processing bodies - P-bodies which concentrate components of RNA decay machinery) or manifestation of stress conditions (stress granules - SGs). Many different types of insults converge into the formation of SGs which is mostly initiated by the phosphorylation of the alpha subunit of the eukaryotic translation initiation factor 2 (eIF2α) at serine 51. eIF2 forms the ternary complex with GTP and Met-tRNA^{iMet}, which provides

the methionine to the 40S subunit of ribosome using energy from the GTP hydrolysis (Kimball, 1999). Phosphorylation of eIF2α blocks the exchange of GDP to GTP bound to eIF2γ by eIF2B. Additionally, this posttranslational modification of eIF2α subunit increases the affinity of eIF2B for eIF2. Both mechanisms contribute to translational suppression. Currently, four kinases are known to control eIF2α through the phosphorylation of Ser51: HRI, GCN2, PERK and PKR. HRI (Heme-regulated eIF2alpha kinase), as its name suggests, responds to the levels of heme molecule (Chen, 2006). GCN2 (general control nonderepressible 2) detects the amino acids deprivation through binding to uncharged tRNAs (Zaborske *et al.*, 2009). The unfolded protein response (UPR) triggers activation of PERK kinase. Finally, PKR (mentioned in the previous paragraphs) detects dsRNA or RNAs with 5' triphosphate and partial secondary structure and elicits translation arrest. In addition, PKR may be activated by a dsRNA-independent mechanism involving PACT (PKR activating protein) (Patel and Sen, 1998; Patel *et al.*, 2000). The limiting amounts of GTP-eIF2 and eIF2B (guanine nucleotide exchange factor) cause accumulation of stalled ribosomes and accessory mRNA interacting proteins. The down-stream events lead to the aggregation of RNA-protein complexes. Several molecules have been demonstrated to be critical for the SGs assembly, maintenance and stability: G3BP1 (GTPase Activating Protein SH3 Domain Binding Protein 1), TIA1, TIAR (T cell restricted intracellular antigen-1, TIA-related protein). An interesting aspect of SGs formation is the involvement of prion-like domains which mediate aggregation as shown for TIA1 (Gilks *et al.*, 2004). Stress granules are highly dynamic structures and serve as a storage point for mRNAs, where the decision about their fate is made. Transcripts may be kept for future translation when stress conditions cease. Alternatively, the messenger RNA can be sent for

degradation in processing bodies (P-bodies), which are structures associated with SGs. The destiny of captured mRNA may depend on the length, type, strength of external assault and input from signaling pathways (**Figure 1.2**).

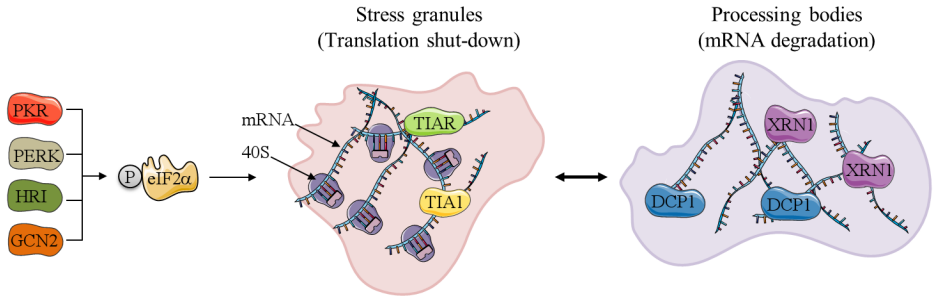


Figure 1.2: Cytoplasmic granules: stress granules and processing bodies. Different stress conditions activate one or more kinases: PKR, GCN2, HRI and PERK which phosphorylate eIF2α. Phosphorylation of eIF2α is responsible for a translational arrest which initiates a cascade of events leading to the formation of stress granules. We depicted only some components of SGs like TIA1, TIAR, G3BP1, 40S ribosomal subunit. Stress granules are linked to processing bodies (P-bodies are also found under normal, rest conditions). P-bodies are primarily involved in the degradation of mRNAs. We presented only two molecules which are known markers for P-bodies: XRN1 (5'-3' exoribonuclease participating in the mRNA decay), DCP1 (decapping protein, component of mRNA decapping complex removing 7-methyl guanine cap from mRNA)

The formation of stress granules is also associated with the translation shut-down during viral infection. A halt of the protein synthesis pathway blocks viral replication and gives cells time to trigger antiviral responses. Thus, viruses evolved many mechanisms to manipulate SGs. Viral infection may result in: the inhibition, transient formation, oscillation (continuous assembly and disassembly) or stable formation of SGs. One of the strategies controlling SGs assembly is the blockage of PKR functions. Influenza A virus (IAV) uses NS1 (nonstructural protein 1) to arrest the SGs assembly. NS1 has an RNA-binding domain that is postulated to compete with PKR for nucleic acids. IAV with NS1 defective in the RNA binding cannot repress stress granules (Khaperskyy *et al.*, 2012). Another scheme counteracting SGs

maintenance involves viral proteases that degrade stress granule components. Currently, only some picornaviruses (i.e. poliovirus) are known to utilize such a mechanism (Tsai and Lloyd, 2014). In the initial stages of poliovirus infection, transient SGs are formed, which are then dissembled due to G3BP1 cleavage at the glutamine residue (Q326) by the viral 3C proteinase. The importance of G3BP1 in the maintenance of the stress granules has been demonstrated in the experiment with the cleavage resistant version of the protein (G3BP1 Q326E) - in these conditions, the stress granules persisted throughout the course of infection (White *et al.*, 2007). Vaccinia virus (VV) which replicates in the cytoplasm was postulated to redirect several stress granule components (i.e. G3BP1, eIF4E, eIF4G, CAPRIN1) to be used in the viral replication factories (RF) (Katsafanas and Moss, 2007). However, in a more recent study involving vaccinia virus without E3L protein (antagonist of PKR), SGs proteins (i.e. TIA1, G3BP1) were detected in the structures surrounding the VV replication factories. The ribosomal subunits (characteristic of bona fide SGs) were not present in these aggregates and it was proposed that these structures are "antiviral" SGs. In contrast, to the previous studies, G3BP1 did not form granular structures within the vaccinia RF which were assumed to be part of "proviral" assemblies. Therefore, further experiments are required to clarify the involvement of SGs components in the vaccinia virus infection (Simpson-Holley *et al.*, 2011).

A growing body of evidence establishes a link between SGs and modulation of the innate immune responses. Stress granules may be regarded as a platform linking stress detection with infection. Indeed, SGs are known to be enriched in proteins involved in innate immunity including PKR, ADAR1 (Okonski and Samuel, 2013), DAI, RLRs (MDA5 - (Langereis *et al.*, 2013), RIG-I (Onomoto *et al.*,

2012; Kuniyoshi *et al.*, 2014), RNase L and OAS. Accumulation of these molecules in SGs may increase the probability of mounting cellular responses as their spatial colocalization would promote the sensing of nucleic acids. In fact, IAV without NS1 promoted IFN-β expression which has been attributed to the stress granules assembly. Knockdown of G3BP1 or knockdown of PKR reduced type I IFNs responses (Onomoto *et al.*, 2012). Nevertheless, it is still a possibility that PKR might directly phosphorylate a component of the interferon pathway and stress granules are not required *per se*. Another study suggested that PKR in complex with CAPRIN1 and G3BP1 is translocated to SGs during stress conditions and this may lead to activation of PKR which further boosts SGs formation and antiviral responses (Reineke *et al.*, 2015). In contrast, MDA5 localization to the stress granules did not influence the production of type I IFNs as shown for PKR deficient HeLa cells or non-phosphorylatable eIF2α MEF cells. Only in the mengovirus L infection, knockdown of PKR resulted in a significant decrease of IFN-β production. Despite the lack of apparent involvement in the IFNs production, SGs formation was correlated with lower viral RNA load (Langereis *et al.*, 2013).

Integrating the knowledge about the battery of mechanisms utilized by viruses to interfere with SGs formation/stability corroborates the notion that SGs have important anti-viral functions. Stress granules constitute a platform that integrates several biological processes. SGs are formed as a result of translational shut-down and therefore, they affect protein production of viral proteins. Additionally, SGs facilitate detection of foreign nucleic acids and help to mount robust innate immunity responses as they are enriched in key elements of the innate immunity.

1.1.5 Imbalances in the nucleic acid detection

Although the ability to trigger cellular responses towards the foreign nucleic acids vastly expands the number of recognized pathogens, it imposes a certain risk to the host, as imbalances in the proper recognition of nucleic acids can lead to autoimmunity. Research concerning the Aicardi–Goutières syndrome (AGS) provided a deeper understanding on how disequilibrium in the detection of nucleic acids contributes to pathology. AGS is a rare autoimmune disease, primarily affecting the brain (a progressive encephalopathy) and the skin (with the characteristic chilblains) (Rice *et al.*, 2007). Currently, mutations in six genes have been linked to AGS (TREX1, ADAR1, SAMHD1, RNASEH2A, RNASEH2B and RNASEH2C). It has been estimated that 90% of AGS patients harbor mutations in one of these six genes, indicating that additional genes can be involved in this disease (Oda *et al.*, 2014). It has been shown that TREX1 (3' repair exonuclease, DNase III) assures that DNA from endogenous retroelements and spurious replication elements do not accumulate in the cytoplasm (Yang *et al.*, 2007; Stetson *et al.*, 2008). An excess of these byproducts leads to overproduction of type I interferons (through STING-dependent pathways). Intriguingly, the double knockout mice, lacking TREX and its signaling mediator STING, were rescued from premature mortality and tissue degeneration, highlighting the importance of the cytoplasmic PRRs in the pathology (Gall *et al.*, 2012). Mutations in ADAR1 (double-stranded RNA-specific adenosine deaminase) have also been implicated in AGS. It is shown that disease-causing mutations are found in the catalytic domain and Z-DNA binding domains of ADAR1 (Rice *et al.*, 2012). ADAR1 is an enzyme which deaminates adenines to inosines in double-stranded RNAs (dsRNAs) including host-derived ones. Through this enzymatic reaction, ADAR1 marks self-dsRNA from being

a potential ligand for cytoplasmic RNA sensors (more detailed description in **Section 1.3.1**).

1.1.6 Strategies to distinguish between host and pathogen nucleic acids

In order to react to potential threats, cells need to distinguish between their own and foreign nucleic acids. The mechanisms of detection of non-self-nucleic acids can be based on either sequence or structure recognition (i.e. Sox2 in neutrophils recognizes specific sequences in bacterial genomes (Xia *et al.*, 2015), TLR13 interacts with bacterial 23S rRNA (Oldenburg *et al.*, 2012)). Alternatively, a lack of specific host modifications (examples given below) or the presence of unusual pathogen marks (i.e. DNA phosphorothioation in bacteria) may be recognized by the host. Theoretically, the sequence/structure specific detection of nucleic acids requires a whole set of sensors which would respond to particular types of pathogen. The system based on the modifications of the self-DNA/RNA seems to be more versatile. TLR9 was the first PRR discovered and it was shown to detect CpG DNA and trigger production of type I interferons and proinflammatory cytokines by the activation of IRF7 and NF-κB. Here a distinction is based on the fact that CpG dinucleotides are methylated and underrepresented in the mammalian genomes. In contrast, DNA of viral or bacterial origin has mostly unmethylated CpG dinucleotides which can elicit immune responses (Hemmi *et al.*, 2000). Another sensor IFI16 has been proposed to detect the foreign DNA because viral nucleic acids are not associated with nucleosomes and/or nucleosome patterning is less compact than that of the host (Orzalli *et al.*, 2013; Morrone *et al.*, 2013). IFI16 is postulated to detect viruses replicating in the nucleus: herpesviruses (Dell'Oste *et al.*, 2015) and most likely

papillomaviruses (Lo Cigno *et al.*, 2015). IFI16 has been shown to assemble oligomers on the naked DNA in a cooperative manner which suggests that it creates on-DNA signaling scaffold depending on the length of the nucleic acid molecule. Nevertheless, the issue concerning the mechanism of the distinction between the host and viral nucleic acids by IFI16 requires further investigation.

RNA may also be modified, and some modifications render it non-immunogenic. As mentioned in the previous paragraphs, RIG-I detects 5'-triphosphate and activates production of type I IFN. Host mRNA is not detected by this sensor because its 5'-end is capped by the modification of ribose with 7-methyl guanosine (m^7G) and 2'-O-methyl group (Devarkar *et al.*, 2016). Thus, many viruses hijack the host system to get 5'-cap to mimic host mRNA (Decroly *et al.*, 2012). Other host RNAs (tRNA, most of rRNA, some snRNA) are targeted for the cleavage and contain monophosphate which does not elicit immune responses (Nallagatla *et al.*, 2008).

It is well established that cytoplasmic dsRNA has A strong immunostimulatory potential. Thus, the compartmentalization of the eukaryotic cell helps in the discrimination between host and foreign dsRNA. A-to-I RNA editing emerges as an attractive model explaining how cells cope with potentially self-activating dsRNA regions of mRNA. During transcription, ADAR1 marks regions of transcripts forming extensive double-stranded structures with inosine. Such modifications destabilize the base-paired regions and eliminate the recognition by the cytoplasmic sensors (Mannion *et al.*, 2014; Liddicoat *et al.*, 2015). In addition, inosine-containing RNAs can suppress interferon induction and apoptosis (Vitali and Scadden, 2010). It has also been shown that transcripts with inosine might be targeted for degradation by EndoV and Tudor-SN complex (Scadden, 2005; Morita *et al.*, 2013).

Currently, we know several types of RNA modifications which may play a role in the innate immunity. It was observed that *in vitro* synthesized RNA with incorporated modified nucleotides (i.e. pseudouridine, 2'-O-methylated, m^5C, m^6A) can weaken the innate immune responses (Karikó *et al.*, 2005; Anderson *et al.*, 2011). Hence, it is possible that the general *modus operandi* of the detection of foreign nucleic acids is based on the fact that they lack certain modifications.

Beyond proteins that recognize B-DNA and A-RNA several sensor proteins (or their inhibitors) bear domains which can interact and stabilize an alternative conformation of nucleic acids. Intriguingly, a subset of host proteins involved in the interferon pathways (ADAR1, DAI and PKZ) or viral evasion proteins (E3L, ORF112), contains Z-DNA/Z-RNA binding domains (Zαs). These structural motifs have the ability to recognize the left-handed conformation of nucleic acids (either dsRNA or dsDNA). Z-DNA/Z-RNA is a high-energy conformation which is formed by purine-pyrimidine repeats, preferably CpG. *In vivo*, transcription can generate the negative supercoiling which provides the energy to stabilize the Z-conformation. Having proteins that recognize this unusual nucleic acid conformation and viral counter partners implicated in the inhibition of antiviral responses suggest that even a transient conformation of nucleic acid may serve as an input for the innate immunity. Yet, we await the evidence that Zα domains can recognize the left-handed conformation in the context of infection. We also do not know what would be the source generating the negative supercoiling leading to the left-handed conformation in infection. In which context the Z-form of DNA or RNA in the cytoplasm can be interpreted as a threat and result in cellular responses? The following sections would provide the summary of the knowledge about the left-handed conformation

and proteins containing Zα domains.

1.2 The left-handed nucleic acids

In 1953, Watson and Crick presented a model for the DNA helical structure. In this model two anti-parallel strands create a right-handed helix. Bases are located inside the helix with the phosphate backbone on the outer surface. Their predictions were based on limited amount of data from fiber diffraction experiments (Wilkins and Randall, 1953; Watson and Crick, 1953; Franklin and Gosling, 1953). At that time obtaining crystals was out of reach as their formation requires a pure, homogeneous sample and DNA synthesis methods were not yet developed. The first crystal structure of DNA was obtained 25 years later in the laboratory of Alexander Rich. Surprisingly, the self-complementary $d(CG)_3$ oligonucleotides showed a left-handed helical conformation (Wang *et al.*, 1979). This conformation was named Z-DNA due to a characteristic zig-zag phosphate backbone. Although it was known by spectrometry experiments that B-DNA is the most prevalent form, Z-DNA attracted a lot of interest mixed with skepticism regarding its biological relevance.

B- and Z-DNA are double helices and based on the Watson-Crick's base pairing. Significant differences exist between these conformations. First, the sense of the helix for Z-DNA is left-handed, in contrast to the right-handed B-DNA. The Z-DNA helix has a smaller diameter than B-DNA (1.8 nm compared to 2 nm). A full turn of the helix requires 12 bp for Z-DNA while B-DNA needs only 10.5 bp. Additionally, Z-DNA has the major groove that is almost flat. The only Z-DNA groove (corresponding to the minor groove of B-DNA) is deep and narrow. A characteristic feature of the left-handed helix is a repeated pattern of syn and anti conformation

of the bases for each dinucleotide. The syn and anti conformations refer to the rotation around the N-glycosidic bond joining the sugar with the base (**Figure 1.3**).

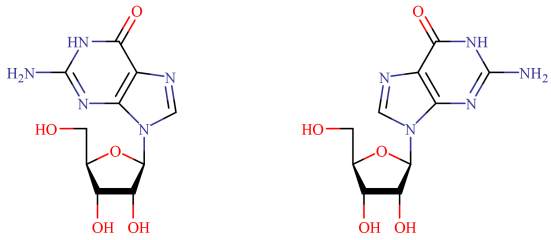


Figure 1.3: Syn and anti conformation of guanosine. These conformers are formed by the rotation around the glycosidic bond.

1.2.1 Z-DNA/Z-RNA

The discovery of the Z-DNA extended a list of known conformations of DNA. Now we know that double-stranded nucleic acids can adopt one of three major conformations: A, B, and Z. *In vivo* dsDNA is usually found as a right-handed B-form. In contrast, the right-handed A-form is readily adopted by dsRNA. For simplicity Z-DNA/Z-RNA and B-DNA will be compared (see **Figure 1.4** and **Table 1.1**).

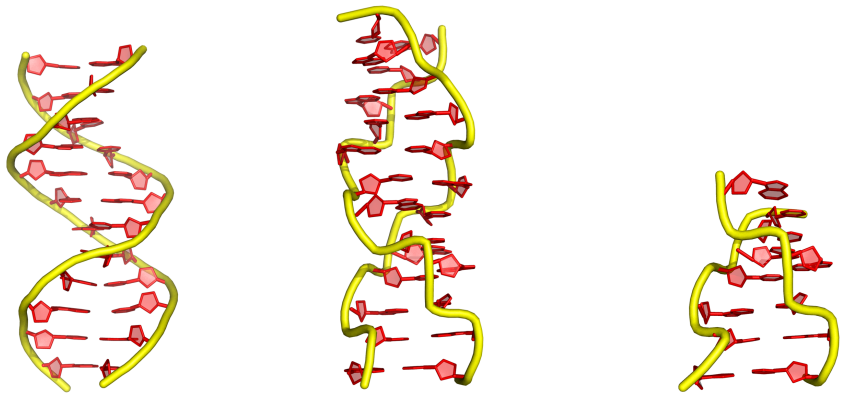


Figure 1.4: Comparison of B-DNA, Z-DNA and Z-RNA. The 12 base pair B-DNA (left, PDB: 1HQ7) and Z-DNA (middle, PDB: 4LB5). In the right side, 6bp Z-RNA helix is showed for comparison (PDB: 2GXB, excluding unpaired region).

Table 1.1: Comparison of B-DNA, Z-DNA and Z-RNA features (Rich *et al.*, 1984).

Feature	B-DNA	Z-DNA	Z-RNA
Helix	right-handed	left-handed	left-handed
Residues per turn	10.5	12	12
Glycosidic torsion angle			
purine	anti	syn	syn
pyrimidine	anti	anti	anti
Diameter (Å)	20	18	18

Purines adopt the syn conformation more readily than pyrimidines (Haschemeyer and Rich, 1967) and as a consequence, the repeating pattern of purine-pyrimidine is the most favorable for the Z-DNA formation. Moreover, in Z-DNA sugar puckers are found to be C3'-endo for purines and C2'-endo for pyrimidines while the canonical B-DNA has C2'-endo sugar pucker for all bases. The sugar pucker nomenclature describes the relative position of the given carbon relative to the plane of the sugar (**Figure 1.5**).

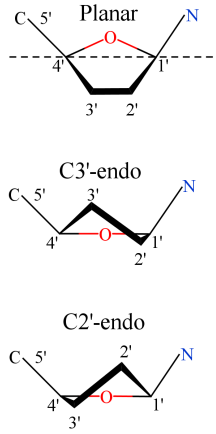


Figure 1.5: C2'-endo and C3'-endo sugar puckering.

Interestingly, although Z-DNA forms a more rigid structure than B-DNA, it can also exhibit structural variation. Analysis of different crystal structures suggested that Z-DNA can adopt slightly different conformations of the phosphate group of GpC step. These phosphates face the minor groove of the helix or point away from it. These Z-DNA forms were named Z_I and Z_{II}, respectively (Wang *et al.*, 1981).

The secondary structure of macromolecules can be evaluated using polarized light by the means of circular dichroism (CD) spectroscopy. Soon after the discovery of Z-DNA, it was realized that inversion of circular dichroism spectra of CG repeats in the high salt conditions presented by Pohl and Jovin in early 70' (Pohl and Jovin, 1972) reflects a transition from B-DNA to Z-DNA conformation (Thamann *et al.*, 1981). Therefore, one of the commonly used methods to distinguish the left-handed from the right-handed conformation is to compare the CD spectra. **Figure 1.6** depicts the idealized CD spectra of B- and Z-DNA. B-DNA spectrum (blue solid line) shows positive ellipticity between 265-300 nm and a negative peak at around 250 nm. In the presence of high salt (i.e. 3-4 M NaCl) $(CG)_n$ spectrum is "inverted" which is a signature of the left-handed conformation. Z-DNA (red dashed line) has a positive peak at around 265 nm and negative around 290 nm.

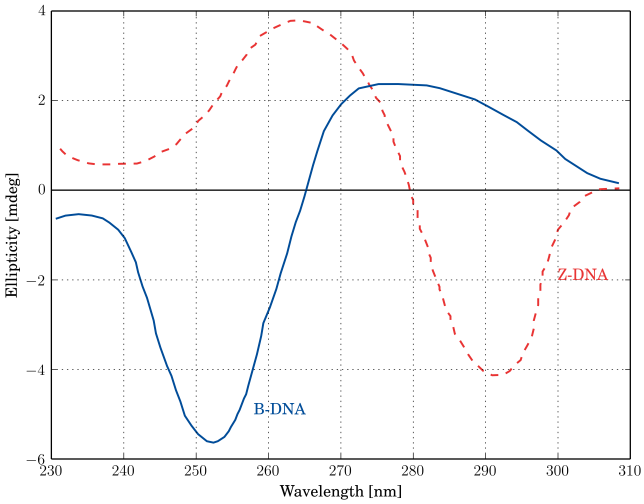


Figure 1.6: Idealized circular dichroism spectrum for B- and Z-DNA. B-DNA (blue solid line) is characterized by positive ellipticity between 265-300 nm. B-form has an ellipticity negative peak at around 250 nm. Z-DNA (red dashed line) has a positive peak at around 265 nm and a negative peak at around 290 nm.

Sequence prerequisites and stability of Z-DNA

After the discovery of Z-DNA great interest was shown in the study of this unusual conformation in order to identify its role in biological processes i.e. gene regulation. Experiments have been devised to assess the propensity of a given purine-pyrimidine repeat to form Z-DNA, either by inducing the B-Z transition using high salt or in the context of negatively supercoiled circular plasmid. It has been established that $(CG)_n$ forms Z-DNA most readily, followed by $(TG)_n$ and $(AT)_n$ having the lowest Z-forming propensity (Jovin *et al.*, 1983). The energy required for the B-Z transition for $(AT)_n$ is higher than for other pyrimidine-purine repeats. It has been postulated that base-base and base-phosphate interactions play a significant role in the low propensity of these repeats for B-Z transition (Dang *et al.*, 1990). In addition, it has been observed that $(AT)_n$ may adopt another non-B-DNA structure: cruciform (Greaves *et al.*, 1985).

Changes of sugar puckering from C2' to C3' contribute to the shortening of the distance between phosphate groups in the Z-conformer. As phosphates are closer in space in Z-DNA than in B-DNA this leads to electrostatic repulsion of these residues. *In vitro*, the unfavorable charge distribution of the left-handed conformation can be shielded by the presence of high salt concentration. The cation valence is an important parameter in the efficiency to neutralize negative charges of the phosphate groups. Monovalent ions (i.e. Na^+, K^+) require high concentration (2-4 M) to be able to stabilize the Z-conformation, whereas divalent ions (i.e. Mg^{2+}, Co^{2+}) would perform equally well at lower concentrations. Cobalt hexamine chloride $Co[NH_3]_6^{3+}$ is an extremely potent stabilizer of Z-DNA and only micromolar concentrations are required for the induction. Moreover, the naturally occurring polyamines i.e. spermine and spermidine also have the capacity to stabilize the left-handed conformation (Jovin *et al.*, 1987).

Stabilization of the Z-conformation is not possible under the physiological salt concentrations. Nevertheless, the observation that chemical modifications of the bases (especially the biologically relevant cytosine methylation) promote Z-formation at lower salt concentrations restored a belief that Z-DNA can have an impact on biology (Behe and Felsenfeld, 1981; Fujii *et al.*, 1982; Zacharias *et al.*, 1990). However, over the years, experimental evidence accumulated that negative supercoiling can be the primary stabilizing force of the Z-conformer *in vivo* either in the context of plasmid supercoiling or transcription (discussed in **Paragraph 1.2.3 and 1.2.3**).

Z-RNA

A few years after the discovery of Z-DNA, studies using circular dichroism, NMR and absorbance spectroscopy provided evidence to support the Z-RNA formation (Hall *et al.*, 1984). It has been demonstrated that A-RNA can flip to Z-RNA (A-Z transition) *in vitro* but it requires higher salt concentrations and temperatures (above 35 °C) than for the formation of the left-handed DNA. This suggested that the energy associated with a transition to the Z-conformation is higher for RNA than for DNA. Z-DNA/Z-RNA binding domains called Zα domains (refer to **Section 1.3**) are great tools to study the Z-conformation at low salt conditions. First, it has been shown that these domains bind Z-RNA (Brown *et al.*, 2000) and this enabled the crystallization and structure determination of the RNA in the left-handed conformation (Placido *et al.*, 2007). The structure revealed many similarities between Z-RNA and Z-DNA, not only concerning helical twist angles but also preferences for sugar puckering (guanosines C3'-endo, cytidines C2'-endo). One of the earlier NMR studies of Z-RNA also indicated that Z-RNA is structurally similar to Z-DNA (Davis *et al.*, 1990). Z-DNA and Z-RNA, despite sharing many similarities, differ in the solvation pattern. Z-RNA has 2'-OH groups of cytosine which are bridged by sodium ions. On the other hand, an NMR structure of Z-RNA in high salt conditions was different from Z-DNA or Z-RNA complexed with the Zα domain in respect to the sugar puckering (guanosines adopt C4'-exo and cytosines are found in C2'-endo, C3'-exo, C1'-exo configurations) (Popenda *et al.*, 2004). Based on these finding it was postulated that at least two forms of Z-RNA exist: Z_D and Z_R (Trulson *et al.*, 1987), where the first resembles Z-DNA, and the latter is observed only when the solvation pattern is changed by high-salt.

1.2.2 The B-Z transition and the formation of B-Z and Z-Z junctions

Over the years, several groups have tried to understand the principles governing the formation of the left-handed conformation. The first experiments indicated that conversion between the right-handed and the left-handed DNA helix (B-Z transition) is governed by entropy as enthalpy was close to zero (Pohl and Jovin, 1972). Following studies contradicted this result as they attributed a significant enthalpy contribution (Chaires and Sturtevant, 1986). Many studies aimed at elucidating the molecular mechanism underlying the B-Z transition. Several models have been proposed for the B-Z transition and they can be classified into three categories (Fuertes *et al.*, 2006):

1. base-pair opening:
   (a) without intermediate:
      - The all-or-none model (Pohl and Jovin, 1972),
      - The Wang model (Wang *et al.*, 1979),
      - The solitary excitation model (Zang and Olson, 1987),
      - The helix-coil transition model (Walker and Aboul-ela, 1988),
      - The thermal fluctuation model (Chen and Prohofsky, 1993),

   (b) with intermediate:
      - B*(C)-DNA intermediate model (Goto, 1984),
      - The zipper model (Peck and Wang, 1983),
      - The stretched intermediate model (Lim and Feng, 2005),

2. base-pair rotation (no opening):
   (a) without intermediate:
      - The Olson model (Olson *et al.*, 1983),
      - The Harvey model (Harvey, 1983),
      - The Ansevin-Wang model (Ansevin and Wang, 1990),

   (b) with intermediate:

- The Saenger-Heinemann model (Saenger and Heinemann, 1989),

3. salt-related models:
   (a) The unified B-Z transition model (Guéron *et al.*, 2000),
   (b) The empirical salt-threshold model (Fuertes *et al.*, 2001).

One class of models assumed that the B-Z transition introduces dramatic changes of the helix and requires melting of hydrogen bonds between base pairs (requirement for the base-pair opening). An alternative mechanism postulated that the transition does not involve opening of the base pairs as the B-conformation is flexible enough. The mechanistic explanations favoring base-pair rotation as a driver of B-Z transition were not able to justify the sensitivity of the B-Z junction to single-strand specific nucleases. Both categories of models can be subdivided into groups including (or not including) intermediates. Salt related models (not connected with the two previous) are related to the salt/ion requirements for the B-Z transition and they do not provide any mechanistic insights.

Currently, the zipper model is assumed to be a simple but robust explanation, in agreement with many experimental data for the B-Z transition (see **Figure 1.7**). Although the zipper model accounts for the cooperativity of the B-Z transition (Peck and Wang, 1983), it does not provide a detailed description of structural or dynamic properties of the B-Z transition. According to this model, the B-Z transition can be divided into two phases: nucleation and propagation. The formation of two B-Z junctions is the first step which has the highest energy requirement. The following stages involve the extension of Z-DNA which moves the junction in the opposite directions along the DNA helix until no further sequence may acquire the Z-conformation.

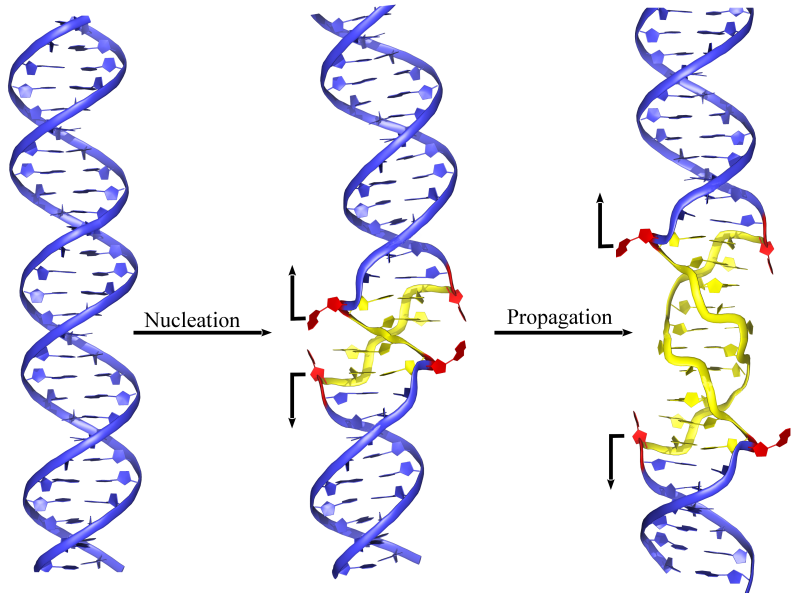


Figure 1.7: A schematic representation of the zipper model for the B-Z transition. The Z-DNA formation can be viewed as two-step process which begins with the creation of the B-Z junctions (nucleation), followed by the propagation stage (the B-Z junctions migrate in the opposite directions). The B-Z junctions are depicted in red, B-DNA in blue, Z-DNA in yellow.

Boundaries of the Z-conformation

In the genomes, perfect purine-pyrimidine repeats are rare as they may contribute to genomic instability (Wang *et al.*, 2006). Moreover, the Z-conformation would always be formed in the context of a B-DNA neighborhood - a segment of Z-DNA would form two B-Z junctions at its ends. Repeats of purine-pyrimidine tracts can also be interrupted by base insertions or deletions that change the phase of the left-handed helices leading to the creation of a Z-Z junction. Structural studies of the boundaries between segments of the left-handed and right-handed (or other portion of Z-DNA) helices were facilitated by the discovery of domains which can stabilize the Z-conformation: Zα domains (see **Section 1.4**). The structures of B-Z and Z-Z junction will be discussed in the following sections.

B-Z junction

The left-handed conformation of DNA is highly energetic, transient and *in vivo* stabilized by the negative supercoiling which can be generated by active transcription and chromatin remodeling. The genomic purine-pyrimidine repeats with a propensity to form Z-DNA are surrounded by sequences which cannot accommodate the Z-conformation. Therefore, whenever B-DNA makes a transition to the left-handed conformation two B-Z junctions are created. In the early years of research on the left-handed conformation, it has been noticed that the B-Z junctions have particular properties. Interestingly, it was observed that nuclease S1 specific for the single-stranded nucleic acids cleaves in the proximity of B-Z junction. Therefore, the region between B-DNA and Z-DNA may have single-stranded properties (Singleton *et al.*, 1982, 1983, 1984). It was also demonstrated that the cleavage of the BamHI restriction site placed in the B-Z junction is reduced to less than 20% of plasmid with low helical density (Singleton *et al.*, 1983). Altogether, these results reinforced the notion that the B-Z junction has a specific structure. Only in 2005 the crystal structure of the B-Z junction was experimentally determined (Ha *et al.*, 2005). In this structure, the left-handed conformation was stabilized by Zα domain from ADAR1. The base pair at the junction was an A:T and these two bases were found to be extruded from the double helix (**Figure 1.8**). This is in agreement with previous Raman spectroscopy studies which suggested that this junction consists of three or fewer base pairs (Dai *et al.*, 1989). This single base pair extrusion is a relatively small structural alteration and overall the helix continuity is maintained. The B-Z junction introduces bending of the two helical axes. There was no interaction of the Zα domains with extruded bases. As the bases of the B-Z junction are exposed to the environment they could be prone to chemical-physical modifications (e.g. oxidation). It is not excluded that enzymes involved in the DNA repair mechanisms or other base modifying enzymes can target these exposed bases (Ha *et al.*, 2005).

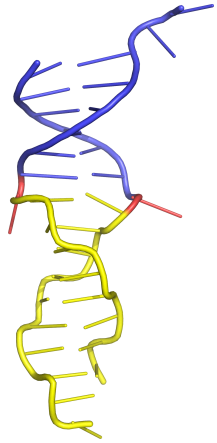


Figure 1.8: The B-Z junction. The extruded junction bases are colored red, Z-DNA is yellow while the B-DNA part is blue. Rendering of structure PDB: 2ACJ, without showing Zα domains.

The Z-Z junction

In most cases, a Z-Z junction would be formed when there are two segments of purine-pyrimidine repeats which are interrupted by a base insertion or deletion. Lack of the dinucleotide pattern continuity results in two left-handed helices out of phase with each other. It is estimated that the Z-Z junction should be as common as the B-Z junction. Runs of CG repeats which are not continuous can be found in CpG islands or Alu elements (Khuu *et al.*, 2007; Su *et al.*, 2014). Moreover, the formation of the Z-Z junction requires less energy than the B-Z junction (Johnston *et al.*, 1991).

In our laboratory, the Zα domain ADAR1 was used to co-crystallize the Z-Z junction in the form of a $(CG)_3A(CG)_3$ oligonucleotide (de Rosa *et al.*, 2010). In one of the crystals of the Z-Z junction, HEPES which was a constituent of the mother liquor, became an intercalating agent at the junction site. Another structure was obtained without this molecule (referred as HEPES free). These structures revealed that the Z-Z junction consists of a single base pair. It was also observed that the DNA is kinked at the Z-Z junction and a cavity is created at the junction site that potentially could be a target for intercalating agents (similar to the observed for the HEPES molecule) (Yang and Wang, 1997). In contrast

to the B-Z junction, the bases at the Z-Z junction are not fully extruded (compare **Figure 1.8** and **Figure 1.9**). An interesting feature of HEPES free structure is that one of the bases of the junction (thymidine) adopts two alternative conformations (syn and anti). This can be explained by the lack of the ability to form a proper hydrogen bond with the pairing adenine. The other base of the junction (adenine) is in the syn conformation (de Rosa *et al.*, 2010).

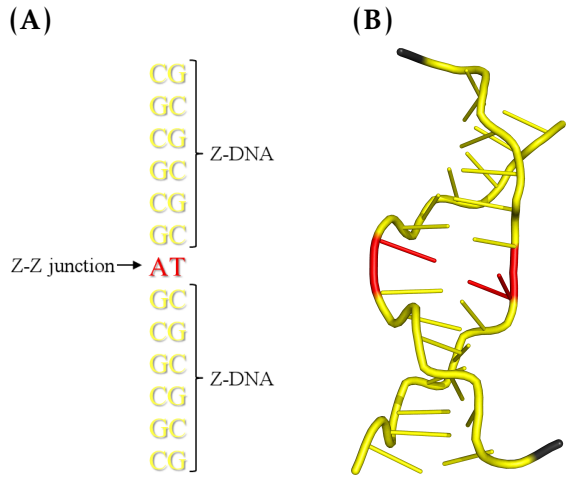


Figure 1.9: The Z-Z junction. (A) The Z-Z junction is formed when the continuity of purine-pyrimidine repeats is interrupted by the insertion or deletion of a base pair. In this example the Z-Z junction is represented by A:T base pair (red). The Z-forming regions are colored in yellow. **(B)** The structural model of the Z-Z junction (as in panel (A)). The bases of Z-Z junction are colored red, Z-DNA is depicted in yellow, an unpaired region in black. Rendering of structure (HEPES free structure) PDB: 3IRQ, without the Zα domains.

1.2.3 Transient *in vivo* formation of the Z-conformation

Chromatin is an extremely dynamic structure composed of DNA wrapped around nucleosomes. Its active rearrangements influence transcription, replication and genome organization. DNA found in the cell is presumed to be in the right-handed form most of the time but regions composed of runs of purine-pyrimidines (able to adopt Z-DNA

conformation) are also prevalent. According to the non-B DNA motif database, the human genome contains more than 400000 sequences with a potential to exist as Z-conformers (Cer *et al.*, 2013). Z-form is a high energy, temporal conformation which requires a stabilizing force. The biophysical characteristics of Z-DNA, such as a rigidity and formation of B-Z/Z-Z junctions, attracted several groups to investigate the formation of Z-DNA in the context of genome stability and chromatin organization.

Z-forming sequences and genomic instability

After the demonstration that Z-DNA can be formed in supercoiled plasmids (Kłysik *et al.*, 1981; Singleton *et al.*, 1982; Nordheim and Rich, 1983; Thomae *et al.*, 1983), it was shown that tracts of CG repeats are not stably propagated in *E.coli*. Inserts with stretches of CG longer than 50 bp were frequently deleted (Kłysik *et al.*, 1982). At that time, authors did not provide any support for the left-handed conformation formed by these fragments. Further evidence supporting Z-DNA formation and genomic instability comes from a similar type of study using pUC plasmids to survey frameshift mutations in a β-galoctosidase complementation assay. Once again, the frequency of frameshift mutations correlated with the length of the CG repeats. It was also noted that $(GT)_{24}$ had frameshift mutations comparable to shorter $(CG)_{12}$. Differences in the energy requirements for $(GC)_n$ and $(GT)_n$ to adopt the left-handed conformation may explain the unequal mutation load (Freund *et al.*, 1989).

Another study assessed the recombination efficiency for non-replicating plasmids in human cells. It has been observed that sequences with $(TG)_{30}$ increase homologous recombination around 20 times compared to the ones without the Z-forming capacity (Wahls *et al.*, 1990). Several reports postulated that sequences with a propensity to form Z-DNA are involved in the chromosomal rearrangements, translocations and breaks in the genes involved in the tumorgenesis like *bcl2* (Adachi and Tsujimoto, 1990), *scl* (Aplan *et al.*, 1992), *c-myc* (Wölfl *et al.*, 1995). To address the role of Z-forming sequences in the genomic instability,

plasmids containing runs of CG repeats or control sequences were evaluated for the mutational patterns in *E.coli* and mammalian cells. It has been discovered that Z-DNA related deletions in mammalian cells are found in a 400 bp region surrounding the potential left-handed region. This contrasts with the observation in *E.coli* where small-scale deletions happen most frequently inside repeats. In addition, large-scale deletions driven by the Z-forming sequences in the mammalian cells seem to be replication independent. This suggests that they may be triggered by the DNA repair machinery (Wang *et al.*, 2006).

In B-DNA bases are found inside the helix, whereas in Z-DNA purines are in syn-conformation and are exposed to the environment (e.g. atoms N7, C8 of guanosine). These residues can potentially be targeted by chemical and physical modifiers (Wang *et al.*, 1979). It has been shown that in a plasmid carrying long Z-forming sequences guanosine was selectively modified by NiCR (nickel reagent). Changes were not observed in shorter sequences of $(CG)_7$. Indeed, NiCR can oxidize N7 atom of guanine residue when exposed to the solvent (Tang *et al.*, 1999). As mentioned before, the structure of the B-Z junction revealed that two nucleotides are extruded, which may make them sensitive to chemical modifications (Ha *et al.*, 2005).

On the other hand, it has been proposed that, in certain conditions, the Z-forming sequences may help to maintain the integrity of the genome by the suppression of other repeat expansions. For instance, the elongation of CCTG repeats in ZNF9 gene is associated with myotonic dystrophy 2 (DM2). It has been found that the first intron of ZNF9 has the Z-forming $(TG)_{15}$ repeat followed by the $(CCTG)_n$ repeats. CCTG repeats may form slipped-strand DNA structures which can lead to deletions or expansions of repeats by replication fork arrest, misalignments, recombinations. In this case, the Z-forming sequence may function as a sink for the negative supercoiling and decrease the probability that the neighboring $(CCTG)_n$ repeat acquires slipped-strand DNA structure. The conservation of the Z-forming element strengthens this notion (Edwards *et al.*, 2009).

Z-DNA in a chromatin context

Methylated CG repeats enabled to test whether the Z-conformation can be incorporated *in vitro* into nucleosomes. Methylation of the 5^{th} atom of cytosines allows the B-Z transition to occur in moderate salt concentrations which do not induce the DNA dissociation from histones. It has been shown that Z-DNA can interact with histones but no nucleosomes are formed (Nickol *et al.*, 1982). On the other hand, the same methylated CG repeats in B-conformation (in the low salt conditions) can generate nucleosomes *in vitro* as judged by the production of 145 bp protected fragment in micrococcal nuclease digestion assay (Nickol *et al.*, 1982). Another group tested nucleosome reconstitution in plasmids with $(GC)_{15}$ repeats. They compared the efficiency of incorporation of CG repeats in the relaxed and supercoiled plasmids. They concluded that repeats in the Z-conformation (in supercoiled plasmid) cannot be incorporated into nucleosomes, whereas CG repeats in the relaxed plasmid (B-conformation) assemble into nucleosomes (Garner and Felsenfeld, 1987). Moreover, the nucleosome-scanning assay indicated that sequences with Z-DNA forming propensity, like $(CG)_9$, have much lower nucleosome occupancy than a random sequence. The experiment was performed in transcription-deficient CYC1 promoter (Wong *et al.*, 2007), and suggests that during the assembly there is a preference to avoid these repeats. Therefore, CG repeats may serve as a linker between two nucleosomes and promote transcription by maintaining open chromatin regions.

A study from 2001 linked chromatin remodeling, Z-DNA formation and the open state of the chromatin that facilitates transcription (Liu *et al.*, 2001). The authors were interested in the genes regulated by the BAF complex. They have found that *CSF1* (colony-stimulating factor 1) gene was one of eighty genes activated by the BAF complex. Using an episomal vector, they were able to prove that the regulation occurs in the context of chromatin (a plasmid that does not incorporate in nucleosomes lacks regulation). The *CSF1* promoter contains a NFI binding site preceded by TG repeats. It was hypothesized that these repeats may be important

for the promoter activity. First, it was demonstrated that repeats are not required for NFI binding (by chromatin immunopercipitation assay). Therefore, if the Z-conformation was important for the regulation of the *CSF1* promoter, CG repeats should activate the promoter to a similar extent as TG repeats whereas random sequences should decrease such activation. Using a luciferase assay, the researchers were able to confirm that CG repeats but not random sequences can substitute TG repeats. *In vivo* Z-DNA formation was verified by Z$\alpha\alpha$FOK digestion (a nuclease domain that contains two Zα ADAR1 domains). This study proposed a model in which NFI binds to the NFI binding site in the *CSF1* promoter. It triggers the recruitment of the BAF complex acting as a chromatin remodeler. The remodeling of chromatin facilitates the Z-DNA formation. The Z-DNA region may help to maintain the chromatin in an open state promoting this way transcription (see **Figure 1.10**) (Liu *et al.*, 2001).

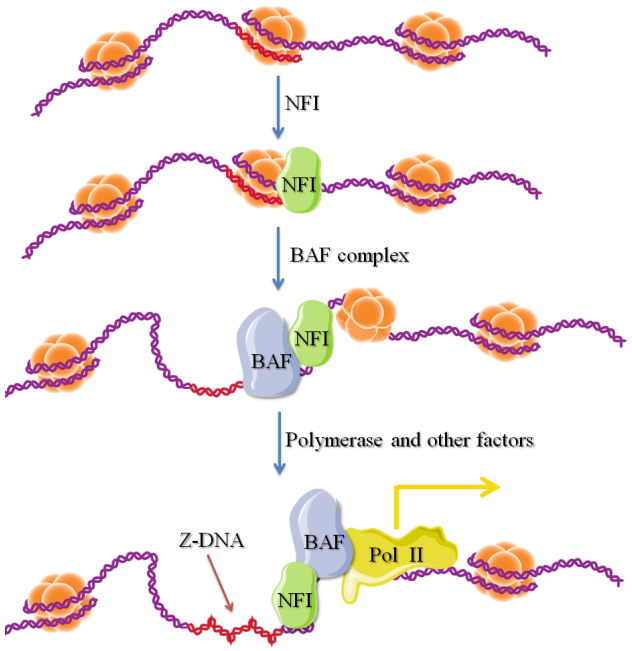


Figure 1.10: Chromatin remodeling and Z-DNA formation. NFI binding recruits BAF complex which facilitates Z-DNA formation. Scheme created based on a model from(Liu *et al.*, 2001).

These results are in agreement with a molecular (Kollman *et al.*, 1982) and a static/dynamic light scattering (Thomas and Bloomfield, 1983)

studies indicating that the Z-conformation is less flexible than B-DNA. An irregular placement of the charges in the zig-zag phosphate backbone can also contribute to the Z-DNA exclusion from nucleosomes facilitating transcription.

Z-conformation and transcription

The *in vitro* conversion between the right-handed and the left-handed helix requires much higher salt concentrations than those found *in vivo*. Early experiments demonstrated that CG repeats in the negatively supercoiled plasmids can acquire the left-handed conformation (Azorin *et al.*, 1983). Transcription emerged as a biologically relevant source of negative supercoiling which can stabilize the left-handed conformation. It has been proposed that during transcription the polymerase can create positive supercoiling in front and negative supercoiling behind it. According to the postulated twin transcriptional-loop model, the polymerase complex goes through DNA and does not rotate around it. As the ends of the nucleic acid are not free, the force is generated when the polymerase is plowing through DNA. This would change the supercoiling of the system (Liu and Wang, 1987).

Several studies provided evidence that the Z-DNA formation is linked to transcriptional activation. The formation of Z-DNA upstream of the promoter may create a nucleosome free-zone exposing important promoter sequences (e.g. TATA box) (Oh *et al.*, 2002). Another supporting evidence for Z-DNA involvement in transcription comes from the analysis of episomal plasmids in yeast *Saccharomyces cerevisiae* (Wong *et al.*, 2007). The authors compared the effects of $(CG)_9$, $(TA)_9$ and C_9G_9 repeats on the transcriptional activation in a β-galactosidase assay. It has been observed that only the $(CG)_9$ repeats strongly activated gene expression in a promoter-dependent manner. Wong and colleagues did not detect an increase in the transcription using $(TA)_9$ and C_9G_9. This suggests that only repeats with a propensity to form Z-DNA but not alternating pyrimidine-purine of $(TA)_9$ or CG content (C_9G_9) are responsible for this

effect. Interestingly, the distance from the TATA box was identified as an important parameter in the transcriptional activation. The increase in the distance between Z-forming sequence and TATA box decreases the activation. One of the explanations for these results can be a requirement of an open chromatin state to facilitate the assembly of the transcription machinery.

The genomic distribution of Z-forming sequences has been evaluated using computational approaches. Z-Hunt software (followed by Z-Hunt II) calculates the propensity of a sequence to undergo the B-Z transition based on the thermodynamic information (Ho *et al.*, 1986; Schroth *et al.*, 1992). This program (employing a modified two-state zipper model) determines the Z-DNA forming potential of an input sequence. First, the software minimizes the total energy required for the B-Z transition (as an effect it maximizes the altering syn-anti conformation). Then, each segment is evaluated based on its ability to transit from B- to Z-DNA in the context of circular plasmid by applying a statistical-mechanical model (the zipper model) (Peck and Wang, 1983). This program has been used to predict sequences with a potential to form Z-DNA in the human genome. In the first study one million base pairs of the human genome were analyzed (Schroth *et al.*, 1992). A few years later, Z-Hunt II was used to predict Z-DNA of the human chromosome 22 (Champ *et al.*, 2004). Both studies provided evidence for a non-random distribution of sequences that may form Z-DNA. It has been observed that the transcription start sites are enriched in sequences with altering purine-pyrimidine repeats. The latter study also assessed the co-occurrence of sequences with a Z-DNA forming propensity with NFI binding sites. It has been discovered that there is a class of genes with Z-forming sequences followed by NFI binding sites in the transcription start sites (TSS). It suggests that these genes may be regulated in a manner similar to *CSF1* gene (compare **Section 1.2.3**, **Figure 1.10**). According to the phylogenomic analysis, metazoans may have two types of Z-forming elements in their promoters. One category is located upstream of the transcription start sites (TSS) and evolved independently of GC and CpG content (composed of TG

repeats). This class of potential Z-DNA regions (ZDRs) is enriched in the NFI binding sites. Second, less frequent, CG-rich class of potential Z-forming sequences is found downstream of TSS (Khuu *et al.*, 2007).

The beauty of the biological systems lies in its dichotomy - the same molecule or state can have two outcomes depending on the conditions and other cues. Similarly, the involvement of Z-forming sequences in transcriptional gene regulation may have two different outcomes. In general, it is accepted that ZDRs would facilitate transcription but in the case of β-maj globin, nucleolin and ADAM-12 genes regions with potential Z-DNA forming regions are linked to the suppression of the gene expression (Gilmour *et al.*, 1984; Rothenburg *et al.*, 2001; Ray *et al.*, 2011). More recently, it has been demonstrated that the negative regulation of ADAM-12 expression requires MeCP2 (methyl-CpG binding protein) and NF1C/NF1X (nuclear factor 1 family factors) (Ray *et al.*, 2013). The details of how Z-DNA and epigenetic regulators are integrated into the suppression of transcription require further investigation.

1.2.4 *In vivo* Z-RNA detection

A limited number of studies explored the formation of Z-RNA *in vivo*. In one the reports, antibodies against Z-RNA were generated and used to probe a potential left-handed RNA in fixed cells of protozoans (Zarling *et al.*, 1987). Authors validated the specificity of anti-Z-RNA antibodies by the radio-immunoassay. It has been demonstrated that the staining with anti-Z-RNA antibody vanishes if samples were treated with RNase A or T1, but not with DNase I. Moreover, in the competition assay the immunofluorescence signal was blocked by Z-RNA but not A-RNA, ssRNA, Z-DNA, which further suggested that, at least in fixed cells, the Z-conformation of RNA is present.

1.3 Proteins binding Z-DNA/Z-RNA

As mentioned earlier, after several years of research on a role of Z-DNA many scientists believed that the Z-conformation was just a curiosity, interesting from the biophysical point of view but without any well-established function in biology. On the other hand, if the Z-conformation was involved in biological processes it was reasonable to assume that cells should possess protein(s) which would be able to interact with Z-DNA. As a result, many attempts have been made to find such Z-DNA binding proteins. One of the first proteins which were shown to bind Z-DNA were antibodies generated against brominated CG repeats a modification that permanently forces the Z-conformation (Lafer *et al.*, 1981). Interestingly, anti-Z-DNA antibodies were also detected in the blood samples of patients suffering from systemic lupus erythematosus (Lafer *et al.*, 1983). Antibodies against the Z-conformations were extensively used to probe many cell types for the presence of Z-DNA/Z-RNA. Over the 80's and early 90's many reports communicated a discovery of proteins with preferential binding to Z-DNA from many different sources i.e. *Drosophila melanogaster* (Nordheim *et al.*, 1982), wheat germ (Lafer *et al.*, 1985), *Saccharomyces cerevisiae* (Zhang *et al.*, 1992) and bull testis (Gut *et al.*, 1987). Some of these findings have been challenged by additional experiments, as for example proteins from the bull testis appeared to recognize bromination of CG repeats, not the conformation (Christen *et al.*, 1990; Rohner *et al.*, 1990).

The problems with the detection of bona fide Z-DNA/Z-RNA binding proteins may lie in the transient nature of this conformation. A DNA probe in the stable left-handed conformation was needed. Alan Herbert in Rich's laboratory responded to that demand developing a robust and sensitive method to isolate Z-DNA binding proteins based on the band-shift assay using linear CG repeats with 5-bromodeoxycytosines as Z-DNA target. The brominated $(CG)_n$ self-complementary duplex could adopt the Z-conformation in 10 mM $MgCl_2$. Additionally, this strategy allowed ^{32}P labeling and the execution of competition assays (Herbert and Rich, 1993). This experimental approach with 20000 fold excess of B-DNA led

to the identification of ADAR1 (double-stranded RNA-specific adenosine deaminase) - as a protein capable of specific interaction with the left-handed DNA (Herbert *et al.*, 1995). Later, this Z-DNA binding activity was mapped to a conserved N-terminal domain of ADAR1 and this domain has been named Zα domain (Herbert *et al.*, 1997).

After deciphering that Zα of ADAR1 is a domain responsible for the interaction with Z-DNA, its sequence has been used to search for other proteins containing similar motifs. Up to date, five proteins containing Zα domains have been identified (see **Figure 1.11**).

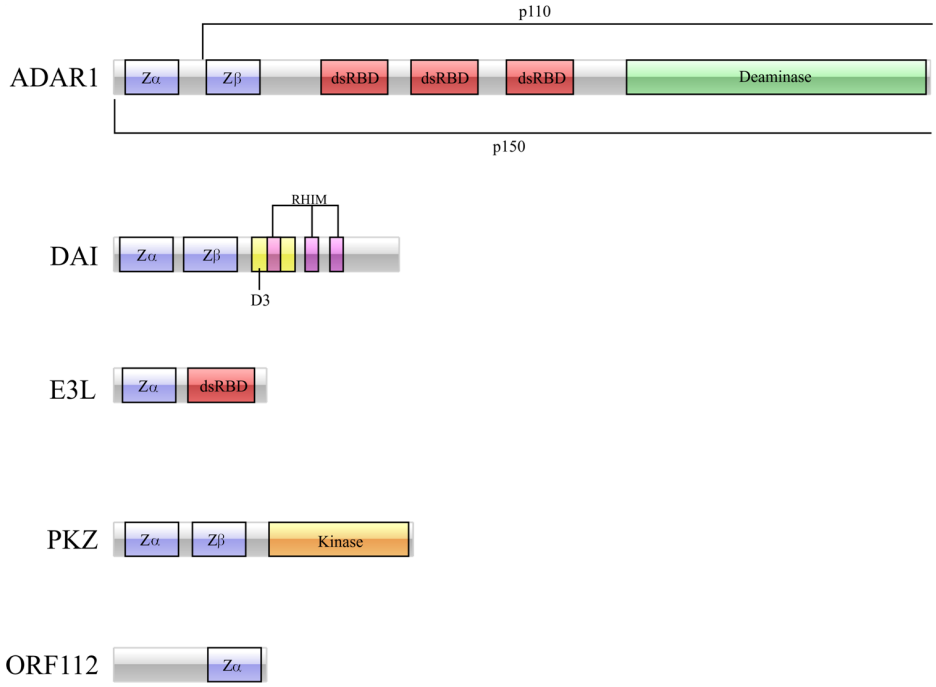


Figure 1.11: Proteins with Zα domains. We depicted domain composition of proteins containing Zα domains. If a protein contains more than one motif with Z-DNA/Z-RNA binding domain signature due to historical reasons the first domain is named Zα and the other Zβ. Unfortunately, this nomenclature lacks consistency because Zβ ADAR1 was not shown to interact with Z-DNA/Z-RNA but Zβ DAI can bind Z-conformation.

Interestingly, all Zα containing proteins are involved in the interferon-mediated innate immunity (ADAR1, DAI and PKZ) or viral proteins counteracting these responses (Zα of E3L and ORF112) (Athanasiadis,

2012). The proteins with Z-DNA binding domains and structures of the different Zα domains are briefly discussed in the following paragraphs.

1.3.1 ADAR1

Experiments designed to clarify unsuccessful gene silencing using antisense RNA in developing embryos of *Xenopus laevis* led to the discovery of a dsRNA unwinding activity (Bass and Weintraub, 1987; Rebagliati and Melton, 1987). Follow-up studies demonstrated that the unwinding of dsRNA occurs due to a covalent modification of dsRNA duplexes. Adenines were detected to undergo conversion and they migrated as inosines in thin-layer chromatography (TLC) (Bass and Weintraub, 1988). The class of proteins which has the ability to perform a catalytic deamination of adenosine was named DRADA (Double-stranded dsRNA adenosine deaminase) and was later renamed to ADAR (Adenosine Deaminase Acting on dsRNA). Now it is known that ADARs are involved in RNA editing - the post-transcriptional modification of pre-mRNA. They carry out deamination of adenosines to inosines in the dsRNA (see **Figure 1.12**). The resulting inosine has properties similar to guanosine (G) (creates base-pairing with cytosine) and is recognized as G by most of the cellular processes.

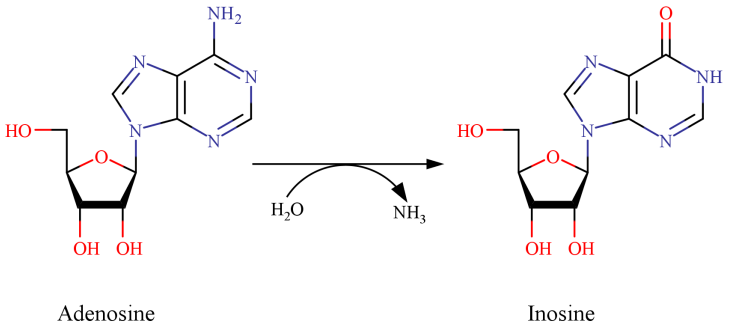


Figure 1.12: Deamination of adenosine produces inosine. ADARs perform a hydrolytic deamination of adenosine resulting in inosine. Inosine is recognized by cellular machinery as guanosine which can result in recoding of amino acids, creation/removal of splicing site and affect miRNA processing.

ADARs do not exhibit any strong sequence specificity apart from a preference for the 5' base neighboring adenosine (Polson and Bass, 1994). Instead, the secondary structure of RNA dictates the degree and the pattern of editing. Up to 50% of adenosines can be edited in perfect duplexes (promiscuous editing), whereas RNA structures with bulges and loops can be targeted with a high specificity on few or a single site (specific editing) (Nishikura, 2010). As inosine is read as guanosine, editing can impact many different processes, leading to recoding of amino acid sequences (Burns *et al.*, 1997) (including destruction of stop codons (Polson *et al.*, 1996)), removal or creation of splice sites (Laurencikiene *et al.*, 2006), modification of miRNA binding sites or modulation of miRNA specificity (Luciano *et al.*, 2004). Initially, most of the discovered RNA editing concerned changes in neurotransmitter receptors i.e. GluR-2,-5,-6 (Sommer *et al.*, 1991; Higuchi *et al.*, 1993) and serotonin-2C (5-HT2C) receptor (Burns *et al.*, 1997). Thus for long, it was widely believed that A-to-I RNA editing is a mechanism for proteome diversification in the animal nervous system. However, analysis of cDNAs (or by next-generation sequencing data) revealed a dramatic level of A-to-I sites in inverted repetitive sequences embedded in mRNAs (Alu elements) which can form nearly perfect dsRNAs (Athanasiadis *et al.*, 2004; Blow *et al.*, 2004; Kim *et al.*, 2004; Levanon *et al.*, 2004; Bazak *et al.*, 2014; Daniel *et al.*, 2014) suggesting alternative functional roles of the modification.

It has been found that the human genome encodes three different ADARs genes (*adar1*, *adar2*, *adar3*). Interestingly, ADAR3 is not known to exhibit any enzymatic activity. ADAR1 has two promoters: a constitutive, driving the expression of the short isoform (p110) and an inducible (controlled by interferons type I and II), producing the long form (p150) (see **Figure 1.11**). These two isoforms also differ in their cellular localization: p110 is mostly nuclear while p150 shuttles between the nucleus and the cytoplasm. Human ADAR1 is a modular protein and it contains dsRBDs (three) and a catalytic domain like other ADARs. In addition, ADAR1 was the first reported protein to contain specific Z-DNA binding activity (Herbert *et al.*, 1995) and later this activity was

attributed to an N-terminal Zα domain (Herbert *et al.*, 1997). ADAR1 contains a second Zα-like domain at the beginning of the constitutive form. Historically, the first domain was named Zα and the second Zβ. ADAR1 p150 isoform contains both Zα and Zβ domains while p110 is shorter and only includes Zβ domain (see **Figure 1.11**). Among domains that belong to the Zα family - the Zβ ADAR1 is the only that does not interact with nucleic acids as it has a mutation in the residue critical for the interaction (see below). Unfortunately, this nomenclature imposes some problems as Zβ domains of other proteins or even the Zβ of zebrafish ADAR1 do bind Z-DNA (Kim *et al.*, 2003)). Nevertheless, Zβ ADAR1 preserves typical Zα fold suggesting that this domain adapted a different function (Athanasiadis *et al.*, 2005).

Over many years, it has been documented that enzymes belonging do ADAR family are found in almost all animals (Grice and Degnan, 2015). However, a Zα containing ADAR1 was thought to be present only from sea anemones onwards but recently the genes encoding ADAR1-like proteins were detected in the sponge genomes (i.e. *Oscarella carmela*, *Amphimedon queenslandica*) (Jin *et al.*, 2009).

ADAR1 and immunity

The eukaryotic cell is composed of functional compartments. As I discussed earlier (see the section: **Detection of DNA**), the separation between the nucleus and cytoplasm may contribute in the distinction of self and foreign nucleic acids by the innate immunity receptors. For instance, dsRNA found in the cytoplasm is regarded as a danger signal and a sign of a viral invasion and results in activation of anti-viral pathways. Antiviral responses against dsRNA can be classified into two major categories: the RNA interference pathway (invertebrates and plants) and interferon responses (vertebrates) (Karpala *et al.*, 2005). ADARs (including ADAR1) can promiscuously edit dsRNA and are involved in the modification of dsRNA intermediates of viruses which was thought to have an antiviral role. The first report of promiscuous A-to-G changes in viruses concerned vesicular stomatitis virus (O'Hara *et al.*, 1984) but at that time it was left unexplained. A few years later, similar "biased

hypermutation" was detected in measles virus (Cattaneo *et al.*, 1988). Soon after the discovery dsRNA unwinding activity it was proposed that this process can be an explanation for the observed mutational pattern in the viral genomes (Bass *et al.*, 1989). Currently, few more virus genomes are known to be affected by ADARs hyperediting (i.e. HCV (Taylor *et al.*, 2005)). Although hyperediting scrambles genetic information of the virus, it is not fully conclusive whether this alteration has antiviral effects.

ADARs can also perform a site-specific editing of viral genomes. Particularly interesting is the specific editing of the hepatitis delta virus (HDV) genome. In this case, the virus takes advantage of the host machinery to alter stop codon to tryptophan and two variants of the hepatitis delta antigen are produced. This change initiates the transition between the replication and the packaging stage (Wu *et al.*, 1995).

On the contrary of an antiviral role, recent data support a rather proviral role of ADAR1 as a negative regulator of antiviral responses affecting both RNAi and the interferon response. The link between immunity and ADAR1 initially emerged from the ADAR1 knock-out studies. It was demonstrated that the knock-out of ADAR1 in mice is embryonic lethal - embryos die before 12.5 days with defects in liver hematopoiesis and extensive apoptosis (Hartner *et al.*, 2004; Wang *et al.*, 2004). Experiments involving induced ADAR1 gene disruption suggested that this protein is required for the maintenance of hematopoietic stem cells and suppression of interferon responses (Hartner *et al.*, 2009). Additionally, knockout of the interferon-inducible ADAR1 p150 was reported and results indicated that the long ADAR1 isoform is responsible for the observed phenotypes (Ward *et al.*, 2011). Recently, a homozygous knock-in of a catalytically inactive mutant of ADAR1 was shown to recapitulate a null mutant. Moreover, the lethal ADAR1$^{E861A/E861A}$ phenotype was rescued by deletion of MDA5 (one of the key components in the dsRNA recognition) (Liddicoat *et al.*, 2015). These results confirmed the outcomes of the previous studies showing that the null ADAR1 lethality can be circumvented by ablation of MAVS (downstream of the sensors MDA5 and RIG-I) (Mannion *et al.*, 2014). Therefore, ADAR1

editing is critical in the prevention of recognition of self-dsRNA and imbalances in the proper recognition of these nucleic acids lead to severe disease. As mentioned in the previous section, mutations found in the catalytic and Zα domain of human ADAR1 are one of the underlying causes of autoimmune Aicardi-Goutières syndrome (Rice *et al.*, 2012). Taken together, ADAR1 is one of the key players in the suppression of the interferon responses. ADAR1 p110 edits dsRNA within mRNAs in the nucleus. This modification would assure that cellular dsRNA would not activate cytoplasmic sensors. The long ADAR1 p150 isoform is only expressed in the presence of interferons and has the ability to shuttle between the nucleus and cytoplasm. The cytoplasmic presence of p150 would act as a negative feedback loop to control an excessive interferon production by sequestering and editing either self or viral dsRNA (**Figure 1.13**).

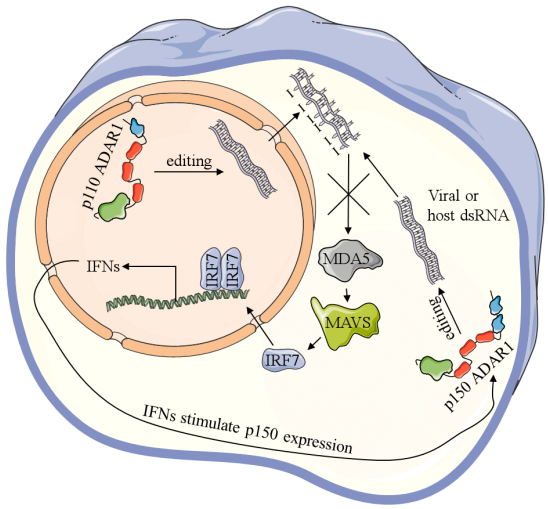


Figure 1.13: ADAR1 function as a negative regulator of interferon responses. ADAR1 is expressed as two isoforms: p110 (constitutive form, localized in the nucleus) and interferon-induced p150. The longer p150 ADAR1 differs from p110 ADAR1 by the N-terminal Zα domain. ADAR1 contains Zα domains (blue), three dsRBD domains (red) and a catalytic domain (green). Edited cellular transcripts cannot elicit immune responses through the MDA5 pathway preventing the production of interferons by self-RNA. During viral infection or uncontrolled release of dsRNA, interferons are produced which in turn activate expression of interferon-induced genes (including p150 ADAR1). p150 ADAR1 shuttles between nucleus and cytoplasm and edits/sequester dsRNA which down-regulates interferon responses and restores homeostasis.

1.3.2 DAI

The second identified protein with Zα domains was DAI. The DNA-dependent activator of IFN-regulatory factors (DAI, alternative names: DLM-1, ZBP1) was identified as a gene highly up-regulated in the ascites tumor stromal cells or after a stimulation of peritoneal macrophages with IFN-γ or LPS (Fu *et al.*, 1999). Several years later, it has been demonstrated that targeting mRNA of DAI with RNA interference weakens the interferon responses to the DNA of various sources. Additionally, the expression of transfected DAI constructs enhanced production of type I interferon. Further experiments indicated that DAI exerts its function by activating of IFN-regulatory factor 3 (IRF3) with the aid of TBK1 (Takaoka *et al.*, 2007). DAI is a modular protein with two N-terminal Zα domains. More recently, it has been uncovered that DAI possesses three RIP homotypic interaction motifs (RHIM motifs) and the first of them overlaps with a region involved in B-DNA binding (see **Figure 1.11**, (Kaiser *et al.*, 2008)). The investigation of the deletion mutants indicated that B-DNA binding (of poly(dA-dT)·(dT-dA)) is primarily carried out by a region called D3. However, the full activation of the interferon production required also Zα domains (Wang *et al.*, 2008). Following studies have uncovered that IRF3 pathway is not the only one activated by nucleic acid bound DAI. It has also been shown that DAI can trigger the NF-κB pathway and as a result, inflammatory cytokines (i.e. IL-6, TNFα) are produced. This activation was found to be dependent on the interaction with RIP1 (the adaptor receptor-interacting protein kinase). The association between DAI and RIP1 is mediated by the RHIM motifs (Kaiser *et al.*, 2008). Furthermore, the interaction of RHIM motifs of DAI with RIP3 has been implicated in the programmed necrosis induced by murine cytomegalovirus (MCMV) (Upton *et al.*, 2012). However, it is still not clear whether this process is triggered by nucleic acid recognition (**Figure 1.14**).

DAI was the first cytoplasmic DNA sensor discovered but later more receptors of this type were identified indicating a redundancy of cytoplasmic DNA sensing (Atianand and Fitzgerald, 2013). Therefore, it

is not surprising that in some cell lines (e.g. mouse embryonic fibroblasts - MEFs) suppression of DAI does not cause to the inhibition of interferon production (Wang *et al.*, 2008; Lippmann *et al.*, 2008). Similarly, MEFs and bone marrow dendritic cells derived from DAI knockout mouse showed that this protein is not required for the production of IFNβ (Ishii *et al.*, 2008) after stimulation with B-DNA. Thus, the role of DAI in the detection of nucleic acids appears to be cell type specific or that the right type of nucleic acids for its activation is not yet truly identified.

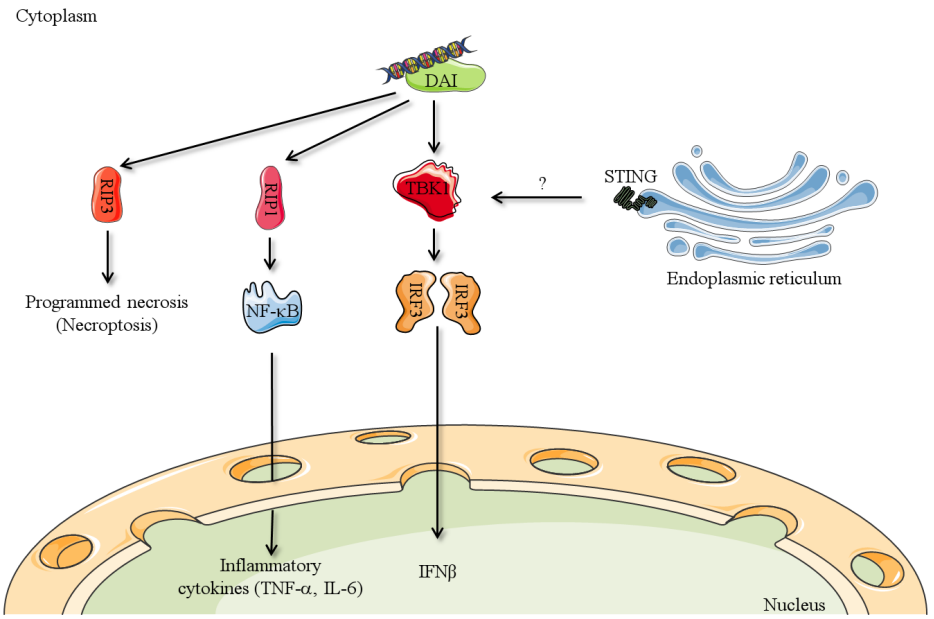


Figure 1.14: Pathways activated by DAI. Upon binding to nucleic acids DAI can engage in different pathways. It has been shown that it can trigger production of IFNβ through TBK1/IRF3 pathway. Moreover, it can interact with RIP1/RIP3 activating either necroptosis or production of cytokines via NF-κB.

1.3.3 E3L

PKR (double-stranded RNA-activated protein kinase) is well known for its antiviral properties. After recognition of double-stranded RNA, PKR forms a dimer and subsequently autophosphorylates. The activated protein phosphorylates eIF2α, which shuts down the translation. The

suppression of protein translation would decrease the amount of proteins available for virus replication limiting viral spread. It has been shown that there are many viral proteins dedicated to the evasion of the PKR action. One of them is E3L - a protein found in poxviruses. In 1984, it was shown that vaccinia virus E3L is an inhibitor of PKR which exerts its action upon binding to the dsRNA (to suppress recognition of dsRNA and circumvent the interferon production) (Whitaker-Dowling and Youngner, 1984; Watson *et al.*, 1991). It has been confirmed that E3L contains a dsRNA binding domain (dsRBD) (Chang and Jacobs, 1993; Yuwen *et al.*, 1993; Chang *et al.*, 1992) but also unexpectedly a Zα domain (Herbert *et al.*, 1997) (see **Figure 1.11**). In cell culture, vaccinia virus (vv) with E3L lacking Zα can replicate similarly to WT virus (Chang *et al.*, 1995; Shors *et al.*, 1997). Thus, the Zα domain is not required for E3L function in this system. However and in opposition to the cell culture experiments, both domains (Zα and dsRBD) of E3L are required for full pathogenicity in a mouse model (Brandt and Jacobs, 2001). Interestingly, Zα domains from ADAR1 or DAI can replace Zα E3L and restore the pathogenicity of vaccinia virus. Chimeric protein with Zβ ADAR1 (which is not known to bind the Z-conformation) instead of Zα E3L, however, does not restore pathogenicity. A mutation that restores Z-DNA/Z-RNA binding in Zβ ADAR1 (I335Y) in the context of chimeric E3L produces indeed a pathogenic virus (Kim *et al.*, 2003). One possible mechanism of the Zα E3L action is a competition for nucleic acids with DAI. Similar, to the mechanism that has been postulated for the dsRBD of E3L which would shield dsRNA from activating PKR or OAS (Willis *et al.*, 2011; Silverman, 2007).

1.3.4 PKZ

As I discussed earlier, in vertebrates, PKR is one of the key players involved in recognition of the foreign dsRNA. This protein is the central node of the translation shut-down and interferon responses. Interestingly, in some fish species a PKR-like protein has been found, which contains

two Zα domains instead of dsRNA binding domains (Hu *et al.*, 2004; Rothenburg *et al.*, 2005). Due to the presence of the Z-DNA/Z-RNA binding domains it has been named PKZ. Analysis of the genomes of these fish species revealed that they contain both PKR and PKZ (Rothenburg *et al.*, 2008). It has been demonstrated that PKZ can phosphorylate (although to a lower extent than PKR) eIF2α (Liu *et al.*, 2011) and, at least, *in vitro*, it can be activated by the interaction with Z-forming CpG repeats (Bergan *et al.*, 2008). Moreover, PKZ has been proposed to be involved in the promotion of the apoptosis (Wu *et al.*, 2015). Therefore, PKR and PKZ in fish seem to have some non-overlapping functions but also act cooperatively to trigger antiviral responses (Liu *et al.*, 2011; Taghavi and Samuel, 2013). It is possible that PKZ would respond to Z-forming dsDNA or dsRNA, whereas PKR would primarily sense the A-form of dsRNA.

1.3.5 ORF112

Cyprinid herpesvirus 3 (CyHV3) is a member of the *Alloherpesviridae* family in the order of *Herpesvirales* (Rakus *et al.*, 2013). Currently, only common and koi carp are known to manifest disease symptoms when infected with CyHV3 but several other fish species are susceptible to its infection (Perelberg *et al.*, 2003). CyHV3 is responsible for huge economic losses and has strong prevalence in the wild (Rakus *et al.*, 2013). Cyprinid herpesviruses have a linear, double-stranded DNA genome which was sequenced several years ago. Genome analyses revealed that at least four genes of CyHV3 share similarities with poxviral genes (Ilouze *et al.*, 2006). Therefore, both poxviruses and alloherpesviruses have a common ancestor or were subjected to a horizontal gene transfer. Using motif-based searches we found that a CyHV3 gene ORF112 encodes for a protein product that contains the Z-DNA/Z-RNA binding motif. Up to now, the only viral protein with a Zα domain was E3L. Thus, ORF112 is an additional protein which parallels poxviruses (E3L) and cyprinid herpesviruses. Unlike poxviral E3L, ORF112 does not contain dsRBD and the N-terminal part of ORF112 is composed of a low complexity region of unknown role.

Intriguingly, ORF112 is encoded in the genome of viruses which infect fish species with PKZ. Thus, ORF112 may function as an antagonist of PKZ in the affected species. It is possible that host and viral Zα domains would compete for nucleic acids. Part of my thesis work was to demonstrate that Zα ORF112 interacts with Z-DNA and that its structure has a typical Zα domain fold and finally clarify the mechanism of DNA/RNA binding for this protein. These findings are discussed in detail in the **Chapter 2** and 3

1.3.6 Proteins with Zα domains localize to stress granules

Different types of cellular stress (including infection) result in a protein translation halt and lead to the formation of cytoplasmic RNP structures known as stress granules. SGs are cytoplasmic assemblies of mRNA and protein complexes (including stalled 40S ribosomes). Multiple proteins are targeted to SGs including several nucleic acids sensors (Onomoto *et al.*, 2012). Proteins with Zα domains under stress conditions also localize to stress granules ((Deigendesch *et al.*, 2006; Ng *et al.*, 2013; Kuś *et al.*, 2015), Gabriel unpublished). In fact, it was demonstrated that Zα domains are sufficient for SGs localization of Zα containing proteins (Ng *et al.*, 2013). Additionally, the association of Zα domains with stress granules requires its Z-DNA/Z-RNA binding activity. It is an important observation because it suggests that these domains interact with nucleic acids not only during infection. As SGs are cytoplasmic, RNA-based structures, Zα domains most likely interact with Z-RNA within such structures. Currently, the nature of the nucleic acids bound by Zα domains in stress granules is not known.

1.4 The Zα domain structure and its interaction with Z-DNA/Z-RNA

In 1999, the structure of the complex of Zα ADAR1 and $T(CG)_3$ was solved and it was the first atomic model of a protein bound to Z-DNA. It constituted an important step in the understanding of the structural details of the Zα interaction with nucleic acids and established a reference to which other structures of Zα can be compared (Schwartz, 1999). This and other structures revealed that Zα domains are representatives of a larger family of the winged-Helix-Turn-Helix (wHTH) domains. Zα domains contain three α-helices and three β-strands. The Zα structure revealed that amino acids that form contacts with Z-DNA come from a $\alpha 3$ helix and a wing region (a loop between $\beta 2$ and $\beta 3$ strands). The triad of amino acids (numbering following Zα ADAR1): Tyr177, Trp195, Asn173 are absolutely critical for the Z-conformation binding and are completely conserved among Zα domains. Mutations in any of these residues abolish the Z-conformation binding. Almost all protein contacts with Z-DNA are formed with the phosphate backbone with the exception of Tyr177 which makes a CH-π bond with guanosine base in the syn conformation (characteristic of Z-form) (see **Figure 1.15**). Later work shows that Zα interacts in a similar manner with Z-RNA providing strong evidence that these domains recognize the shape and not the type of nucleic acid (Placido *et al.*, 2007).

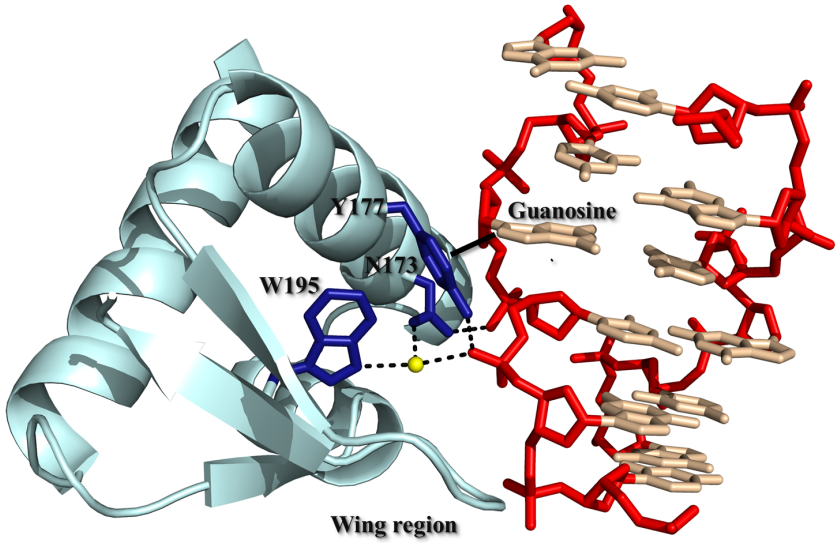


Figure 1.15: Details of interaction between Zα ADAR1 and Z-DNA. Zα ADAR1 is represented as a cyan cartoon. The triad of critical, Z-DNA interacting amino acids is presented as a stick model (dark blue). Z-DNA is shown as a stick model (red and wheat colors correspond to phosphate/sugar backbone and bases, respectively). The hydrogen bonds are drawn as black, dotted lines. The unique, direct interaction between Tyr177 and guanosine in the syn conformation is marked as a solid black line. Rendering of the structure PDB: 1QBJ, chain A.

Most of the crystal structures of Zα complexes with Z-DNA used $T(CG)_3$ as a minimal binding site (for reference **Table 1.2**) and two Zα monomers are found on the opposite side of the duplex without having protein-protein contacts (Schwartz, 1999). However, recent studies with longer Z-DNA oligonucleotides (including the ones presented in this work) uncovered that alternative configurations of the protein on the Z-DNA are possible and Zα monomers can engage in the protein-protein interactions.

Table 1.2: Structural studies of Z-DNA binding domains.

PDB ID	Protein	Method	Organism	Nucleic acid
1QBJ	Zα ADAR1	X-ray	*H. sapiens*	$T(CG)_3$
1J75	Zα DAI	X-ray	*M. musculus*	$T(CG)_3$
1SFU	Zα E3L	X-ray	Yaba virus	$T(CG)_3$
1OYI	Zα E3L	NMR	Vaccinia virus	-
1XMK	Zβ ADAR1	X-ray	*H. sapiens*	-
2ACJ	Zα ADAR1	X-ray	*H. sapiens*	B-Z junction
2GBX	Zα ADAR1	X-ray	*H. sapiens*	RNA T(CG)3
3EYI	Zβ DAI	X-ray	*H. sapiens*	$T(CG)_3$
2L4M	Zβ DAI	NMR	*H. sapiens*	-
3IRQ, 3IRR	Zα ADAR1	X-ray	*H. sapiens*	Z-Z junction
4HOB*	Zα ORF112	X-ray	CyHV-3	-
4LB5, 4LB6	Zα PKZ	X-ray	*D. rerio*	$T(CG)_6$
4KMF	Zα PKZ	X-ray	*C. auratus*	$T(CG)_3$
4WCG*	Zα ORF112	X-ray	CyHV-3	$T(CG)_9$

* Present work.

1.4.1 Other structural/biochemical studies on Zα domains

The interaction of the Z-DNA binding domains of DAI with nucleic acids has been studied in detail on the structural level. Crystal structure of Zα mouse DAI (mDAI) with $T(CG)_3$ has revealed the conserved set of interactions with DNA consisting of the triad of Tyr, Asn, Trp. It has been proven that despite the relatively weak sequence conservation the overall fold is preserved. Notable differences in the recognition of Z-form between Zα mDAI and Zα hADAR1 concern a wing region. Four residues

of the wing region of Zα hADAR1 contact DNA whereas only three from Zα mDAI (Schwartz *et al.*, 2001). Interestingly, X-ray studies of the second Z-DNA binding domain of human DAI (Zβ) have uncovered higher degree of variation in the Z-DNA recognition. The global fold has not changed and the set of three core amino acid interacting with Z-DNA is preserved. In the case of Zβ hDAI, only one residue of the wing is involved in the recognition of Z-DNA. Additional contacts between protein and DNA are contributed by arginine from the first β-strand. Another difference is found in the α3 recognition helix, where a part of this region adopts 3_{10}-helix conformation (Ha *et al.*, 2008). The unusual binding mode of Zβ DAI has been confirmed by NMR studies. The authors also investigated the free protein and postulated that, in contrast to Zα ADAR1, the Zβ hDAI undergoes conformational changes upon Z-DNA binding. It has been noted that tyrosine (which forms a CH-π bond with a base in the syn conformation) is not prepositioned to bind Z-DNA. Therefore, this residue undergoes a conformational change when the protein is bound to left-handed nucleic acid (Kim *et al.*, 2011).

Although we still lack the knowledge about *in vivo* targets of PKZ, the binding mode of Zα PKZ has provided interesting insights into the understanding how these domains may recognize substrates longer than a minimal $T(CG)_3$ sequence. First, the structure of Zα from *D. rerio* PKZ with $T(CG)_6$ revealed that these domains may engage in the second strand interaction with nucleic acids (until the determination of this structure it was thought that the Zα monomer interacts with only one of the strands) (de Rosa *et al.*, 2013). In another structure of Zα from *Carassius auratus* PKZ (caPKZ), the authors uncovered that a lysine residue in the wing region of Zα caPKZ contributes to the fast B-Z transition, as measured by circular dichroism (Kim *et al.*, 2014).

The binding affinity of vaccinia virus Zα E3L has been compared with a canonical Zα ADAR1. *In vitro* analysis revealed that the affinity of vaccinia virus Zα E3L is significantly lower than that of Zα ADAR1 (Kahmann *et al.*, 2004). Even more interestingly, the CD spectrum inversion typical for a B-Z transition of the CG repeat is not observed

in the presence of Zα E3L (Quyen *et al.*, 2007) (**Figure 1.6**). Based on NMR structure, the critical tyrosine residue (Y48) in the free protein has a different side chain conformation than in complex and it is exposed to the solvent. This difference can be an underlying cause for a lower affinity for nucleic acids compared with other Zαs, as there would be a cost associated with Y48 rearrangement (Kahmann *et al.*, 2004). Another Zα E3L from Yaba-like disease virus was crystallized in complex with $T(CG)_3$ oligonucleotide and its structure was solved (discussed in greater detail in **Chapter 3**) (Ha *et al.*, 2004). Yaba-like disease virus Zα E3L binding to Z-DNA is similar to recognition of the left-handed nucleic acids by Zα ADAR1 (Ha *et al.*, 2004).

1.4.2 The Zα domains binding mechanism

As the Z-conformation is transient, it immediately raises the question whether Zα domains actively induce the B-Z transition or if these domains selectively interact with pre-formed Z-DNA/Z-RNA a result of a B-to-Z equilibrium, shifting this equilibrium towards Z-form. This aspect of Zα biochemistry has been investigated by many different groups with some conflicting results. The NMR structure of free Zα ADAR1 domain revealed that most of the residues (seven out of nine) involved in the interaction with the nucleic acids are in the same conformation as in the DNA-bound form (Schade *et al.*, 1999). This result suggested that Zα does not undergo significant structural changes and has its key residues ready to bind to the Z-conformation. Moreover, the modeling of multiple Zα ADAR1·B-DNA complexes showed that binding to right-handed DNA is disfavored by protein clashes either with a minor or a major DNA groove (Schade *et al.*, 1999). Taken together, these results would advocate for the passive mechanism. Another line of evidence supporting that Zα selects the Z-conformation in the sea of B-DNA comes from single-molecule FRET studies (Bae *et al.*, 2011). Using a methylated $(CG)_6$ repeat and Ni^{2+}, B-Z transition dynamics were evaluated in the presence or absence of Zα ADAR1. Bea *et al.* showed that the kinetics of the B-Z transition is

independent of the protein presence or absence suggesting that Zα traps the pre-formed Z-DNA and does not affect the B-Z transition (Bae *et al.*, 2011). Additionally, Zα domains bind both Z-DNA and Z-RNA, therefore, these motifs would require two distinct mechanisms of transition from B-DNA to Z-DNA and A-RNA to Z-RNA making the passive mechanism far more plausible. However, studies involving measurements of hydrogen exchange rates of imino protons by NMR would be in favor of an active mechanism. It has been hypothesized that Zα ADAR1 (Kang *et al.*, 2009; Seo *et al.*, 2010) and yatapoxvirus E3L (Lee *et al.*, 2010) actively transform B-DNA to Z-DNA by an active-mono mechanism (one monomer is able to perform the transition). Another study from the same group postulated that the second Zα domain of DAI uses an active-di mechanism (Kim *et al.*, 2011). Nevertheless, this experimental approach allows only indirect observation of B-DNA binding by Zα domains and one cannot exclude the passive mechanism (Kang *et al.*, 2009).

1.5 Aims and thesis scope

The main objective of the work presented in this thesis is to gain insights into the Z-DNA binding domains structure and biology. Using homology searches we discovered a new protein with a Zα domain - ORF112 - encoded in the genome of cyprinid herpesvirus 3. In **Chapter 2**, we focused on the structural and biochemical analysis of Zα ORF112. We established that the C-terminal part of ORF112 has the ability to interact with DNA in the left-handed conformation. Moreover, we describe the structure of free Zα ORF112. We confirm that, structurally, Zα ORF112 is similar to other Zα domains (i.e. from ADAR1, DAI and E3L). This structure also uncovers an unusual domain-swapped dimer, suggesting that the wing region of the domain is flexible. Given the fact that ORF112 is a putative inhibitor of the interferon responses in some fish species, we decided to reveal the mode of binding of Zα ORF112 to Z-DNA. **Chapter 3** presents a crystal structure of the Zα ORF112 complexed with $T(CG)_9$. We define the network of interactions between this protein and

DNA. We demonstrate that Zα ORF112 recognizes Z-DNA using the same core mechanism as other Zα domains. Nevertheless, we observe an on-DNA protein dimerization, which we postulate (and provide some experimental evidence) to be important in the stabilization of the Z-conformation. In order to validate these structural observations, we generated the corresponding mutants and we evaluate them in the isothermal titration calorimetry. We find that mutations in the protein-protein interface amino acids decrease the relative affinity of the protein to $T(CG)_6$. Furthermore, we use structural information to infer and discuss the evolutionary origins of Zα ORF112.

Several studies have indicated that Zα domains are targeted to stress granules (SGs) under oxidative stress conditions (Deigendesch *et al.*, 2006; Ng *et al.*, 2013). We and others have observed that mutations of key residues involved in the recognition of Z-DNA/Z-RNA abolish Zα domains colocalization with SGs. This provides an indication that Zα domains may indeed interact with host nucleic acids. Thus, we studied Zα domains from DAI – a protein involved in DNA sensing and necroptosis induction. **Chapter 4** describes a technique (developed by us and named tail-Clip) enabling to capture and sequence host nucleic acids bound to Zα domains. In our analysis, we used stable cell lines expressing either Z$\alpha\beta$ DAI or its mutant (not capable of Z-DNA/Z-RNA binding). We have not yet reached final conclusions about ligands of Zα domains as we cannot attribute with statistical significance any specific sequences to wild-type Z$\alpha\beta$ due to high background levels. Hence, we discuss future developments and improvements which may resolve these issues.

Beyond the Zα domain, ADAR1 has dsRBD domains and catalytic domain which are key for the specificity of the enzyme. This protein is implicated in a plethora of cellular functions i.e. modulation of interferon responses through RNA editing. ADARs distinguish their targets on the basis of the extensive recognition of intramolecular dsRNA structures. To obtain insights and understand the specificity, selectivity, catalytic mechanism and substrate recognition by ADARs we employed biochemical and computational approaches. First, we tried to

characterize ADAR1 catalytic domain (with or without dsRBD domain). **Chapter 5** documents our efforts to establish a bacterial or eukaryotic system for overexpression of the catalytic domain constructs of ADAR1 which were not successful. We then decided to understand how ADAR1 dsRNA binding domains (dsRBD) interact with dsRNA. These domains are thought to play an important role in substrate recognition - their cooperation is probably the key determinant of substrate selection for the deamination. Thus, we attempted to characterize the biochemical properties of dsRBDs of ADAR1 (described in **Chapter 5**). Finally, we complemented our biochemical studies with computational studies of ADAR1 substrates **Chapter 5**. We created a computational pipeline to survey co-occurrence of silent point mutations and RNA editing in the predicted RNA secondary structures. We planned to examine how silent point mutations may modulate the editing levels.

Chapter 6 is intended to provide a summary of the findings included in this thesis and highlight possible future research avenues.

Bibliography

Adachi, M. and Tsujimoto, Y. (1990). Potential Z-DNA elements surround the breakpoints of chromosome translocation within the 5' flanking region of bcl-2 gene. *Oncogene*, 5(11):1653–7. /cited on p. 31/

Alexopoulou, L., Holt, A. C., Medzhitov, R., and Flavell, R. A. (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*, 413(6857):732–8. /cited on p. 5/

Anderson, B. R., Muramatsu, H., Jha, B. K., Silverman, R. H., Weissman, D., and Karikó, K. (2011). Nucleoside modifications in RNA limit activation of 2'-5'-oligoadenylate synthetase and increase resistance to cleavage by RNase L. *Nucleic Acids Res.*, 39(21):9329–38. /cited on p. 17/

Ansevin, A. T. and Wang, A. H. (1990). Evidence for a new Z-type left-handed DNA helix: properties of Z(WC)-DNA. *Nucleic Acids Res.*, 18(20):6119–26. /cited on p. 25/

Aplan, P. D., Raimondi, S. C., and Kirsch, I. R. (1992). Disruption of the SCL gene by a t(1;3) translocation in a patient with T cell acute lymphoblastic leukemia. *J. Exp. Med.*, 176(5):1303–10. /cited on p. 31/

Athanasiadis, A. (2012). Zalpha-domains: at the intersection between RAthanasiadis, A., 2012. Zalpha-domains: at the intersection between RNA editing and innate immunity. Seminars in Cell & Developmental Biology.NA editing and innate immunity. *Semin. Cell Dev. Biol.*, 23(3):257–280. /cited on p. 39/

Athanasiadis, A., Placido, D., Maas, S., Brown, B. A., Lowenhaupt, K., and Rich, A. (2005). The crystal structure of the Zbeta domain of the RNA-editing enzyme ADAR1 reveals distinct conserved surfaces among Z-domains. *J. Mol. Biol.*, 351(3):496–507. /cited on p. 42/

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.*, 2(12). /cited on p. 41/

Atianand, M. K. and Fitzgerald, K. A. (2013). Molecular basis of DNA recognition in the immune system. *J. Immunol.*, 190(5):1911–8. /cited on p. 45/

Azorin, F., Nordheim, A., and Rich, A. (1983). Formation of Z-DNA in negatively supercoiled plasmids is sensitive to small changes in salt concentration within the physiological range. *EMBO J.*, 2(5):649–55. /cited on p. 35/

Bae, S., Kim, D., Kim, K. K., Kim, Y.-G., and Hohng, S. (2011). Intrinsic Z-DNA is stabilized by the conformational selection mechanism of Z-DNA-binding proteins. *J. Am. Chem. Soc.*, 133(4):668–71. /cited on p. 54, 55/

Bass, B. L. and Weintraub, H. (1987). A developmentally regulated activity that unwinds RNA duplexes. *Cell*, 48(4):607–13. /cited on p. 40/

Bass, B. L. and Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, 55(6):1089–98. /cited on p. 40/

Bass, B. L., Weintraub, H., Cattaneo, R., and Billeter, M. A. (1989). Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA. *Cell*, 56(3):331. /cited on p. 43/

Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F. J., Rechavi, G., Li, J. B., Eisenberg, E., and Levanon, E. Y. (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.*, 24(3):365–76. /cited on p. 41/

Behe, M. and Felsenfeld, G. (1981). Effects of methylation on a synthetic polynucleotide: the B–Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc. Natl. Acad. Sci. U. S. A.*, 78(3):1619–23. /cited on p. 23/

Bergan, V., Jagus, R., Lauksund, S., Kileng, O. y., and Robertsen, B. r. (2008). The Atlantic salmon Z-DNA binding protein kinase phosphorylates translation initiation factor 2 alpha and constitutes a unique orthologue to the mammalian dsRNA-activated protein kinase R. *FEBS J.*, 275(1):184–97. /cited on p. 48/

Blow, M., Futreal, P. A., Wooster, R., and Stratton, M. R. (2004). A survey of RNA editing in human brain. *Genome Res.*, 14(12):2379–87. /cited on p. 41/

Botos, I., Liu, L., Wang, Y., Segal, D. M., and Davies, D. R. (2009). The toll-like receptor 3:dsRNA signaling complex. *Biochim. Biophys. Acta*, 1789(9-10):667–74. /cited on p. 5/

Brandt, T. A. and Jacobs, B. L. (2001). Both carboxy- and amino-terminal domains of the vaccinia virus interferon resistance gene, E3L, are required for pathogenesis in a mouse model. *J. Virol.*, 75(2):850–6. /cited on p. 47/

Brown, B. A., Lowenhaupt, K., Wilbert, C. M., Hanlon, E. B., and Rich, A. (2000). The zalpha domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 97(25):13532–6. /cited on p. 24/

Burdette, D. L., Monroe, K. M., Sotelo-Troha, K., Iwig, J. S., Eckert, B., Hyodo, M., Hayakawa, Y., and Vance, R. E. (2011). STING is a direct innate immune sensor of cyclic di-GMP. *Nature*, 478(7370):515–518. /cited on p. 7/

Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E., and Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, 387(6630):303–8. /cited on p. 41/

Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., ter Meulen, V., and Billeter, M. A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell*, 55(2):255–65. /cited on p. 43/

Cer, R. Z., Donohue, D. E., Mudunuri, U. S., Temiz, N. A., Loss, M. A., Starner, N. J., Halusa, G. N., Volfovsky, N., Yi, M., Luke, B. T., Bacolla, A., Collins, J. R., and Stephens, R. M. (2013). Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.*, 41(Database issue):D94–D100. /cited on p. 31/

Chaires, J. B. and Sturtevant, J. M. (1986). Thermodynamics of the B to Z transition in poly(m5dG-dC). *Proc. Natl. Acad. Sci. U. S. A.*, 83(15):5479–83. /cited on p. 25/

Chakrabarti, A., Jha, B. K., and Silverman, R. H. (2011). New insights into the role of RNase L in innate immunity. *J. Interf. Cytokine Res.*, 31(1):49–57. /cited on p. 9/

Champ, P. C., Maurice, S., Vargason, J. M., Camp, T., and Ho, P. S. (2004). Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. *Nucleic Acids Res.*, 32(22):6501–10. /cited on p. 36/

Chang, H., Uribe, L., and Jacobs, B. (1995). Rescue of vaccinia virus lacking the E3L gene by mutants of E3L. *J. Virol.*, 69(10):6605–6608. /cited on p. 47/

Chang, H. W. and Jacobs, B. L. (1993). Identification of a conserved motif that is necessary for binding of the vaccinia virus E3L gene products to double-stranded RNA. *Virology*, 194(2):537–47. /cited on p. 47/

Chang, H. W., Watson, J. C., and Jacobs, B. L. (1992). The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. *Proc. Natl. Acad. Sci. U. S. A.*, 89(11):4825–4829. /cited on p. 47/

Chen, J. (2006). Regulation of protein synthesis by the heme-regulated eIF2 kinase: relevance to anemias. *Blood*, 109(7):2693–2699. /cited on p. 10/

Chen, Y. Z. and Prohofsky, E. W. (1993). Salt dependent premelting base pair opening probabilities of B and Z DNA Poly [d(G-C)] and significance for the B-Z transition. *Biophys. J.*, 64(5):1394–7. /cited on p. 25/

Chiu, Y.-H., MacMillan, J. B., and Chen, Z. J. (2009). RNA Polymerase III Detects Cytosolic DNA and Induces Type I Interferons through the RIG-I Pathway. *Cell*, 138(3):576–591. /cited on p. 9/

Christen, T., Bischoff, M., Hobi, R., and Kuenzle, C. C. (1990). High mobility group proteins 1 and 2 bind preferentially to brominated poly(dG-dC).poly(dG-dC) in the Z-DNA conformation but not to other types of Z-DNA. *FEBS Lett.*, 267(1):139–41. /cited on p. 38/

Civril, F., Deimling, T., de Oliveira Mann, C. C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.-P. (2013). Structural mechanism of cytosolic DNA sensing by cGAS. *Nature*, 498(7454):332–7. /cited on p. 6/

Dai, Z. Y., Thomas, G. A., Evertsz, E., and Peticolas, W. L. (1989). The length of a junction between the B and Z conformations in DNA is three base pairs or less. *Biochemistry*, 28(17):6991–6. /cited on p. 28/

Dang, L. X., Pearlman, D. A., and Kollman, P. A. (1990). Why do A.T base pairs inhibit Z-DNA formation? *Proc. Natl. Acad. Sci. U. S. A.*, 87(12):4630–4. /cited on p. 22/

Daniel, C., Silberberg, G., Behm, M., and Öhman, M. (2014). Alu elements shape the primate transcriptome by cis-regulation of RNA editing. *Genome Biol.*, 15(2):R28. /cited on p. 41/

Davis, P. W., Adamiak, R. W., and Tinoco, I. (1990). Z-RNA: the solution NMR structure of r(CGCGCG). *Biopolymers*, 29(1):109–22. /cited on p. 24/

De Nardo, D. (2015). Toll-like receptors: Activation, signalling and transcriptional modulation. *Cytokine*, 74(2):181–9. /cited on p. 5/

de Rosa, M., de Sanctis, D., Rosario, A. L., Archer, M., Rich, A., Athanasiadis, A., and Carrondo, M. A. (2010). Crystal structure of a junction between two Z-DNA helices. *Proc. Natl. Acad. Sci. U. S. A.*, 107(20):9088–92. /cited on p. 29, 30/

de Rosa, M., Zacarias, S., and Athanasiadis, A. (2013). Structural basis for Z-DNA binding and stabilization by the zebrafish Z-DNA dependent protein kinase PKZ. *Nucleic Acids Res.*, 41(21):9924–33. /cited on p. 53/

Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.*, 10(1):51–65. /cited on p. 16/

Deigendesch, N., Koch-Nolte, F., and Rothenburg, S. (2006). ZBP1 subcellular localization and association with stress granules is controlled by its Z-DNA binding domains. *Nucleic Acids Res.*, 34(18):5007–20. /cited on p. 49, 56/

Dell'Oste, V., Gatti, D., Giorgio, A. G., Gariglio, M., Landolfo, S., and De Andrea, M. (2015). The interferon-inducible DNA-sensor protein IFI16: a key player in the antiviral response. *New Microbiol.*, 38(1):5–20. /cited on p. 15/

Devarkar, S. C., Wang, C., Miller, M. T., Ramanathan, A., Jiang, F., Khan, A. G., Patel, S. S., and Marcotrigiano, J. (2016). Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc. Natl. Acad. Sci. U. S. A.*, pages 1515152113–. /cited on p. 16/

Edwards, S. F., Sirito, M., Krahe, R., and Sinden, R. R. (2009). A Z-DNA sequence reduces slipped-strand structure formation in the myotonic dystrophy type 2 (CCTG) x (CAGG) repeat. *Proc. Natl. Acad. Sci. U. S. A.*, 106(9):3270–5. /cited on p. 32/

Elias, P. M. (2007). The skin barrier as an innate immune element. *Semin. Immunopathol.*, 29(1):3–14. /cited on p. 4/

Franklin, R. E. and Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–1. /cited on p. 18/

Freund, A. M., Bichara, M., and Fuchs, R. P. (1989). Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc. Natl. Acad. Sci. U. S. A.*, 86(19):7465–9. /cited on p. 31/

Fu, Y., Comella, N., Tognazzi, K., Brown, L. F., Dvorak, H. F., and Kocher, O. (1999). Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display. *Gene*, 240(1):157–163. /cited on p. 45/

Fuertes, M. A., Cepeda, V., Alonso, C., and Pérez, J. M. (2006). Molecular mechanisms for the B-Z transition in the example of poly[d(G-C) x d(G-C)] polymers. A critical review. *Chem. Rev.*, 106(6):2045–64. /cited on p. 25/

Fuertes, M. A., Pérez, J. M., González, V. M., and Alonso, C. (2001). A kinetic model for the B-Z transition of poly[d(G-C)].poly[d(G-C)] and poly[d(G-m5C)].poly[d(G-m5C)]. *J. Biol. Inorg. Chem.*, 6(7):675–82. /cited on p. 26/

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H., and Rich, A. (1982). Molecular structure of (m 5 dC-dG) 3 : the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res.*, 10(23):7879–7892. /cited on p. 23/

Gall, A., Treuting, P., Elkon, K. B., Loo, Y.-M., Gale, M., Barber, G. N., and Stetson, D. B. (2012). Autoimmunity initiates in nonhematopoietic cells and progresses via lymphocytes in an interferon-dependent autoimmune disease. *Immunity*, 36(1):120–31. /cited on p. 14/

Garner, M. M. and Felsenfeld, G. (1987). Effect of Z-DNA on nucleosome placement. *J. Mol. Biol.*, 196(3):581–590. /cited on p. 33/

Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoecklin, G., Dember, L. M., and Anderson, P. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol. Biol. Cell*, 15(12):5383–98. /cited on p. 10/

Gilmour, R. S., Spandidos, D. A., Vass, J. K., Gow, J. W., and Paul, J. (1984). A negative regulatory sequence near the mouse beta-maj globin gene

associated with a region of potential Z-DNA. *EMBO J.*, 3(6):1263–72. /cited on p. 37/

Goto, S. (1984). Characterization of intermediate conformational states in the B in equilibrium Z transitions of poly(dG-dC).poly(dG-dC). *Biopolymers*, 23(11 Pt 1):2211–22. /cited on p. 25/

Goubau, D., Deddouche, S., and Reis e Sousa, C. (2013). Cytosolic sensing of viruses. *Immunity*, 38(5):855–69. /cited on p. 9/

Greaves, D. R., Patient, R. K., and Lilley, D. M. (1985). Facile cruciform formation by an (A-T)34 sequence from a Xenopus globin gene. *J. Mol. Biol.*, 185(3):461–478. /cited on p. 22/

Grice, L. F. and Degnan, B. M. (2015). The origin of the ADAR gene family and animal RNA editing. *BMC Evol. Biol.*, 15(1):4. /cited on p. 42/

Guéron, M., Demaret, J., and Filoche, M. (2000). A unified theory of the B-Z transition of DNA in high and low concentrations of multivalent ions. *Biophys. J.*, 78(2):1070–83. /cited on p. 26/

Gut, S. H., Bischoff, M., Hobi, R., and Kuenzle, C. C. (1987). Z-DNA-binding proteins from bull testis. *Nucleic Acids Res.*, 15(23):9691–705. /cited on p. 38/

Ha, S. C., Kim, D., Hwang, H.-Y., Rich, A., Kim, Y.-G., and Kim, K. K. (2008). The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 105(52):20671–6. /cited on p. 53/

Ha, S. C., Lokanath, N. K., Van Quyen, D., Wu, C. A., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K. K. (2004). A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 101(40):14367–72. /cited on p. 54/

Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K. K. (2005). Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature*, 437(7062):1183–6. /cited on p. 28, 32/

Habjan, M. and Pichlmair, A. (2015). Cytoplasmic sensing of viral nucleic acids. *Curr. Opin. Virol.*, 11:31–7. /cited on p. 9/

Hall, K., Cruz, P., Tinoco, I., Jovin, T. M., and van de Sande, J. H. (1984). 'Z-RNA'–a left-handed RNA double helix. *Nature*, 311(5986):584–6. /cited on p. 24/

Hartner, J. C., Schmittwolf, C., Kispert, A., Müller, A. M., Higuchi, M., and Seeburg, P. H. (2004). Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J. Biol. Chem.*, 279(6):4894–902. /cited on p. 43/

Hartner, J. C., Walkley, C. R., Lu, J., and Orkin, S. H. (2009). ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nat. Immunol.*, 10(1):109–15. /cited on p. 43/

Harvey, S. C. (1983). DNA structural dynamics: longitudinal breathing as a possible mechanism for the B in equilibrium Z transition. *Nucleic Acids Res.*, 11(14):4867–78. /cited on p. 25/

Haschemeyer, A. and Rich, A. (1967). Nucleoside conformations: an analysis of steric barriers to rotation about the glycosidic bond. *J. Mol. Biol.*, 27(2):369–384. /cited on p. 20/

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. (2004). Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science (80-.).*, 303(5663):1526–9. /cited on p. 5/

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., and Akira, S. (2000). A Toll-like receptor recognizes bacterial DNA. *Nature*, 408(6813):740–5. /cited on p. 5, 15/

Herbert, A., Alfken, J., Kim, Y.-G., Mian, I. S., Nishikura, K., and Rich, A. (1997). A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc. Natl. Acad. Sci. U. S. A.*, 94(16):8421–8426. /cited on p. 39, 42, 47/

Herbert, A., Lowenhaupt, K., Spitzner, J., and Rich, A. (1995). Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 92(16):7550–4. /cited on p. 39, 41/

Herbert, A. G. and Rich, A. (1993). A method to identify and characterize Z-DNA binding proteins using a linear oligodeoxynucleotide. *Nucleic Acids Res.*, 21(11):2669–72. /cited on p. 38/

Higuchi, M., Single, F. N., Köhler, M., Sommer, B., Sprengel, R., and Seeburg, P. H. (1993). RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell*, 75(7):1361–70. /cited on p. 41/

Ho, P. S., Ellison, M. J., Quigley, G. J., and Rich, A. (1986). A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *EMBO J.*, 5(10):2737–44. /cited on p. 36/

Hu, C.-Y., Zhang, Y.-B., Huang, G.-P., Zhang, Q.-Y., and Gui, J.-F. (2004). Molecular cloning and characterisation of a fish PKR-like gene from cultured CAB cells induced by UV-inactivated virus. *Fish Shellfish Immunol.*, 17(4):353–66. /cited on p. 48/

Ilouze, M., Dishon, A., Kahan, T., and Kotler, M. (2006). Cyprinid herpes virus-3 (CyHV-3) bears genes of genetically distant large DNA viruses. *FEBS Lett.*, 580(18):4473–8. /cited on p. 48/

Ishii, K. J., Coban, C., Kato, H., Takahashi, K., Torii, Y., Takeshita, F., Ludwig, H., Sutter, G., Suzuki, K., Hemmi, H., Sato, S., Yamamoto, M., Uematsu, S., Kawai, T., Takeuchi, O., and Akira, S. (2006). A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nat. Immunol.*, 7(1):40–8. /cited on p. 6/

Ishii, K. J., Kawagoe, T., Koyama, S., Matsui, K., Kumar, H., Kawai, T., Uematsu, S., Takeuchi, O., Takeshita, F., Coban, C., and Akira, S. (2008). TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature*, 451(7179):725–9. /cited on p. 46/

Jin, Y., Zhang, W., and Li, Q. (2009). Origins and evolution of ADAR-mediated RNA editing. *IUBMB Life*, 61(6):572–8. /cited on p. 42/

Johnston, B. H., Quigley, G. J., Ellison, M. J., and Rich, A. (1991). The Z-Z junction: the boundary between two out-of-phase Z-DNA regions. *Biochemistry*, 30(21):5257–63. /cited on p. 29/

Jovin, T. M., McIntosh, L. P., Arndt-Jovin, D. J., Zarling, D. A., Robert-Nicoud, M., van de Sande, J. H., Jorgenson, K. F., and Eckstein, F. (1983). Left-handed DNA: from synthetic polymers to chromosomes. *J. Biomol. Struct. Dyn.*, 1(1):21–57. /cited on p. 22/

Jovin, T. M., Soumpasis, D. M., and McIntosh, L. P. (1987). The Transition Between B-DNA and Z-DNA. *Annu. Rev. Phys. Chem.*, 38(1):521–558. /cited on p. 23/

Kahmann, J. D., Wecking, D. A., Putter, V., Lowenhaupt, K., Kim, Y.-G., Schmieder, P., Oschkinat, H., Rich, A., and Schade, M. (2004). The solution structure of the N-terminal domain of E3L shows a tyrosine conformation that may explain its reduced affinity to Z-DNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.*, 101(9):2712–7. /cited on p. 53, 54/

Kaiser, W. J., Upton, J. W., and Mocarski, E. S. (2008). Receptor-interacting protein homotypic interaction motif-dependent control of NF-kappa B activation via the DNA-dependent activator of IFN regulatory factors. *J. Immunol.*, 181(9):6427–34. /cited on p. 45/

Kang, Y.-M., Bang, J., Lee, E.-H., Ahn, H.-C., Seo, Y.-J., Kim, K. K., Kim, Y.-G., Choi, B.-S., and Lee, J.-H. (2009). NMR spectroscopic elucidation of the B-Z transition of a DNA double helix induced by the Z alpha domain of human ADAR1. *J. Am. Chem. Soc.*, 131(32):11485–91. /cited on p. 55/

Karikó, K., Buckstein, M., Ni, H., and Weissman, D. (2005). Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, 23(2):165–75. /cited on p. 17/

Karpala, A. J., Doran, T. J., and Bean, A. G. D. (2005). Immune responses to dsRNA: implications for gene silencing technologies. *Immunol. Cell Biol.*, 83(3):211–6. /cited on p. 42/

Katsafanas, G. C. and Moss, B. (2007). Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. *Cell Host Microbe*, 2(4):221–8. /cited on p. 12/

Kawai, T. and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.*, 11(5):373–84. /cited on p. 4/

Kerur, N., Veettil, M. V., Sharma-Walia, N., Bottero, V., Sadagopan, S., Otageri, P., and Chandran, B. (2011). IFI16 acts as a nuclear pathogen sensor to induce the inflammasome in response to Kaposi Sarcoma-associated herpesvirus infection. *Cell Host Microbe*, 9(5):363–75. /cited on p. 7/

Khaperskyy, D. A., Hatchette, T. F., and McCormick, C. (2012). Influenza A virus inhibits cytoplasmic stress granule formation. *FASEB J.*, 26(4):1629–39. /cited on p. 11/

Khuu, P., Sandor, M., DeYoung, J., and Ho, P. S. (2007). Phylogenomic analysis of the emergence of GC-rich transcription elements. *Proc. Natl. Acad. Sci. U. S. A.*, 104(42):16528–33. /cited on p. 29, 37/

Kim, D., Hur, J., Park, K., Bae, S., Shin, D., Ha, S. C., Hwang, H.-Y., Hohng, S., Lee, J.-H., Lee, S., Kim, Y.-G., and Kim, K. K. (2014). Distinct Z-DNA

binding mode of a PKR-like protein kinase containing a Z-DNA binding domain (PKZ). *Nucleic Acids Res.*, 42(9):5937–48. /cited on p. 53/

Kim, K., Khayrutdinov, B. I., Lee, C.-K., Cheong, H.-K., Kang, S. W., Park, H., Lee, S., Kim, Y.-G., Jee, J., Rich, A., Kim, K. K., and Jeon, Y. H. (2011). Solution structure of the Zbeta domain of human DNA-dependent activator of IFN-regulatory factors and its binding modes to B- and Z-DNAs. *Proc. Natl. Acad. Sci. U. S. A.*, 108(17):6921–6. /cited on p. 53, 55/

Kim, Y.-G., Lowenhaupt, K., Oh, D.-B., Kim, K. K., and Rich, A. (2004). Evidence that vaccinia virulence factor E3L binds to Z-DNA in vivo: Implications for development of a therapy for poxvirus infection. *Proc. Natl. Acad. Sci. U. S. A.*, 101(6):1514–8. /cited on p. 41/

Kim, Y.-G., Muralinath, M., Brandt, T., Pearcy, M., Hauns, K., Lowenhaupt, K., Jacobs, B. L., and Rich, A. (2003). A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, 100(12):6974–9. /cited on p. 42, 47/

Kimball, S. R. (1999). Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.*, 31(1):25–9. /cited on p. 10/

Kollman, P., Weiner, P., Quigley, G., and Wang, A. (1982). Molecular-mechanical studies of Z-DNA: a comparison of the structural and energetic properties of Z- and B-DNA. *Biopolymers*, 21(10):1945–69. /cited on p. 34/

Kuniyoshi, K., Takeuchi, O., Pandey, S., Satoh, T., Iwasaki, H., Akira, S., and Kawai, T. (2014). Pivotal role of RNA-binding E3 ubiquitin ligase MEX3C in RIG-I-mediated antiviral innate immunity. *Proc. Natl. Acad. Sci. U. S. A.*, 111(15):5646–51. /cited on p. 13/

Kuś, K., Rakus, K., Boutier, M., Tsigkri, T., Gabriel, L., Vanderplasschen, A., and Athanasiadis, A. (2015). The Structure of the Cyprinid herpesvirus 3 ORF112-Zα·Z-DNA Complex Reveals a Mechanism of Nucleic Acids Recognition Conserved with E3L, a Poxvirus Inhibitor of Interferon Response. *J. Biol. Chem.*, 290(52):30713–25. /cited on p. 49/

Kłysik, J., Stirdivant, S. M., Larson, J. E., Hart, P. A., and Wells, R. D. (1981). Left-handed DNA in restriction fragments and a recombinant plasmid. *Nature*, 290(5808):672–7. /cited on p. 31/

Kłysik, J., Stirdivant, S. M., and Wells, R. D. (1982). Left-handed DNA. Cloning, characterization, and instability of inserts containing different

lengths of (dC-dG) in Escherichia coli. *J. Biol. Chem.*, 257(17):10152–8. /cited on p. 31/

Lafer, E. M., Möller, A., Nordheim, A., Stollar, B. D., and Rich, A. (1981). Antibodies specific for left-handed Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 78(6):3546–50. /cited on p. 38/

Lafer, E. M., Sousa, R., Rosen, B., Hsu, A., and Rich, A. (1985). Isolation and characterization of Z-DNA binding proteins from wheat germ. *Biochemistry*, 24(19):5070–5076. /cited on p. 38/

Lafer, E. M., Valle, R. P., Möller, A., Nordheim, A., Schur, P. H., Rich, A., and Stollar, B. D. (1983). Z-DNA-specific antibodies in human systemic lupus erythematosus. *J. Clin. Invest.*, 71(2):314–21. /cited on p. 38/

Langereis, M. A., Feng, Q., and van Kuppeveld, F. J. (2013). MDA5 localizes to stress granules, but this localization is not required for the induction of type I interferon. *J. Virol.*, 87(11):6314–25. /cited on p. 12, 13/

Laurencikiene, J., Källman, A. M., Fong, N., Bentley, D. L., and Ohman, M. (2006). RNA editing and alternative splicing: the importance of co-transcriptional coordination. *EMBO Rep.*, 7(3):303–7. /cited on p. 41/

Lee, M., Kim, S. H., and Hong, S.-C. (2010). Minute negative superhelicity is sufficient to induce the B-Z transition in the presence of low tension. *Proc. Natl. Acad. Sci. U. S. A.*, 107(11):4985–90. /cited on p. 55/

Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z. Y., Shoshan, A., Pollock, S. R., Sztybel, D., Olshansky, M., Rechavi, G., and Jantsch, M. F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.*, 22(8):1001–5. /cited on p. 41/

Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).*, 349(6252):1115–1120. /cited on p. 16, 43/

Lim, W. and Feng, Y. P. (2005). The stretched intermediate model of B-Z DNA transition. *Biophys. J.*, 88(3):1593–607. /cited on p. 25/

Lippmann, J., Rothenburg, S., Deigendesch, N., Eitel, J., Meixenberger, K., van Laak, V., Slevogt, H., N'guessan, P. D., Hippenstiel, S., Chakraborty, T., Flieger, A., Suttorp, N., and Opitz, B. (2008). IFNβ responses induced by intracellular bacteria or cytosolic DNA in different human cells do

not require ZBP1 (DLM-1/DAI). *Cell. Microbiol.*, 10(12):2579–88. /cited on p. 46/

Liu, L. F. and Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U. S. A.*, 84(20):7024–7027. /cited on p. 35/

Liu, R., Liu, H., Chen, X., Kirby, M., Brown, P. O., and Zhao, K. (2001). Regulation of CSF1 Promoter by the SWI/SNF-like BAF Complex. *Cell*, 106(3):309–318. /cited on p. 33, 34/

Liu, T.-K., Zhang, Y.-B., Liu, Y., Sun, F., and Gui, J.-F. (2011). Cooperative roles of fish protein kinase containing Z-DNA binding domains and double-stranded RNA-dependent protein kinase in interferon-mediated antiviral response. *J. Virol.*, 85(23):12769–80. /cited on p. 48/

Lo Cigno, I., De Andrea, M., Borgogna, C., Albertini, S., Landini, M. M., Peretti, A., Johnson, K. E., Chandran, B., Landolfo, S., and Gariglio, M. (2015). The Nuclear DNA Sensor IFI16 Acts as a Restriction Factor for Human Papillomavirus Replication through Epigenetic Modifications of the Viral Promoters. *J. Virol.*, 89(15):7506–20. /cited on p. 16/

Luciano, D. J., Mirsky, H., Vendetti, N. J., and Maas, S. (2004). RNA editing of a miRNA precursor. *RNA*, 10(8):1174–7. /cited on p. 41/

Lund, J. M., Alexopoulou, L., Sato, A., Karow, M., Adams, N. C., Gale, N. W., Iwasaki, A., and Flavell, R. A. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl. Acad. Sci. U. S. A.*, 101(15):5598–603. /cited on p. 5/

Mannion, N. M., Greenwood, S. M., Young, R., Cox, S., Brindle, J., Read, D., Nellå ker, C., Vesely, C., Ponting, C. P., McLaughlin, P. J., Jantsch, M. F., Dorin, J., Adams, I. R., Scadden, A. D. J., Ohman, M., Keegan, L. P., and O'Connell, M. A. (2014). The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.*, 9(4):1482–94. /cited on p. 16, 43/

Morita, Y., Shibutani, T., Nakanishi, N., Nishikura, K., Iwai, S., and Kuraoka, I. (2013). Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nat. Commun.*, 4:2273. /cited on p. 16/

Morrone, S. R., Wang, T., Constantoulakis, L. M., Hooy, R. M., Delannoy, M. J., and Sohn, J. (2013). Cooperative assembly of IFI16 filaments on dsDNA provides insights into host defense strategy. *Proc. Natl. Acad. Sci. U. S. A.*, 111(1):E62–E71. /cited on p. 15/

Nallagatla, S. R., Toroney, R., and Bevilacqua, P. C. (2008). A brilliant disguise for self RNA: 5'-end and internal modifications of primary transcripts suppress elements of innate immunity. *RNA Biol.*, 5(3):140–4. /cited on p. 16/

Ng, S. K., Weissbach, R., Ronson, G. E., and Scadden, A. D. J. (2013). Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules. *Nucleic Acids Res.*, 41(21):9786–99. /cited on p. 49, 56/

Nickol, J., Behe, M., and Felsenfeld, G. (1982). Effect of the B–Z transition in poly(dG-m5dC) . poly(dG-m5dC) on nucleosome formation. *Proc. Natl. Acad. Sci. U. S. A.*, 79(6):1771–5. /cited on p. 33/

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, 79:321–49. /cited on p. 41/

Nordheim, A. and Rich, A. (1983). The sequence (dC-dA)n X (dG-dT)n forms left-handed Z-DNA in negatively supercoiled plasmids. *Proc. Natl. Acad. Sci. U. S. A.*, 80(7):1821–5. /cited on p. 31/

Nordheim, A., Tesser, P., Azorin, F., Kwon, Y. H., Moller, A., and Rich, A. (1982). Isolation of Drosophila proteins that bind selectively to left-handed Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 79(24):7729–7733. /cited on p. 38/

Oda, H., Nakagawa, K., Abe, J., Awaya, T., Funabiki, M., Hijikata, A., Nishikomori, R., Funatsuka, M., Ohshima, Y., Sugawara, Y., Yasumi, T., Kato, H., Shirai, T., Ohara, O., Fujita, T., and Heike, T. (2014). Aicardi-Goutières syndrome is caused by IFIH1 mutations. *Am. J. Hum. Genet.*, 95(1):121–5. /cited on p. 14/

Oh, D.-B., Kim, Y.-G., and Rich, A. (2002). Z-DNA-binding proteins can act as potent effectors of gene expression in vivo. *Proc. Natl. Acad. Sci. U. S. A.*, 99(26):16666–71. /cited on p. 35/

O'Hara, P. J., Nichol, S. T., Horodyski, F. M., and Holland, J. J. (1984). Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell*, 36(4):915–24. /cited on p. 42/

Okonski, K. M. and Samuel, C. E. (2013). Stress granule formation induced by measles virus is protein kinase PKR dependent and impaired by RNA adenosine deaminase ADAR1. *J. Virol.*, 87(2):756–66. /cited on p. 12/

Oldenburg, M., Krüger, A., Ferstl, R., Kaufmann, A., Nees, G., Sigmund, A., Bathke, B., Lauterbach, H., Suter, M., Dreher, S., Koedel, U., Akira, S., Kawai, T., Buer, J., Wagner, H., Bauer, S., Hochrein, H., and Kirschning, C. J. (2012). TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance-forming modification. *Science (80-.).*, 337(6098):1111–5. /cited on p. 15/

Olson, W. K., Srinivasan, A. R., Marky, N. L., and Balaji, V. N. (1983). Theoretical probes of DNA conformation examining the B leads to Z conformational transition. *Cold Spring Harb. Symp. Quant. Biol.*, 47 Pt 1:229–41. /cited on p. 25/

Onomoto, K., Jogi, M., Yoo, J.-S., Narita, R., Morimoto, S., Takemura, A., Sambhara, S., Kawaguchi, A., Osari, S., Nagata, K., Matsumiya, T., Namiki, H., Yoneyama, M., and Fujita, T. (2012). Critical role of an antiviral stress granule containing RIG-I and PKR in viral detection and innate immunity. *PLoS One*, 7(8):e43031. /cited on p. 12, 13, 49/

Orzalli, M. H., Conwell, S. E., Berrios, C., DeCaprio, J. A., and Knipe, D. M. (2013). Nuclear interferon-inducible protein 16 promotes silencing of herpesviral and transfected DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 110(47):E4492–501. /cited on p. 15/

Patel, C. V., Handy, I., Goldsmith, T., and Patel, R. C. (2000). PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR. *J. Biol. Chem.*, 275(48):37993–8. /cited on p. 10/

Patel, R. C. and Sen, G. C. (1998). PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J.*, 17(15):4379–90. /cited on p. 10/

Peck, L. J. and Wang, J. C. (1983). Energetics of B-to-Z transition in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 80(20):6206–10. /cited on p. 25, 26, 36/

Perelberg, A., Smirnov, M., Hutoran, M., Diamant, A., Bejerano, Y., and Kotler, M. (2003). Epidemilogical Description Of A New Viral Disease Afflicting Cultured Cyprinus Carpio In Israel. *Isr. J. Aquac. - Bamidgeh*, 55(1):5–12. /cited on p. 48/

Placido, D., Brown, B. A., Lowenhaupt, K., Rich, A., and Athanasiadis, A. (2007). A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure*, 15(4):395–404. /cited on p. 24, 50/

Pohl, F. M. and Jovin, T. M. (1972). Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.*, 67(3):375–396. /cited on p. 21, 25/

Polson, A. G. and Bass, B. L. (1994). Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.*, 13(23):5701–11. /cited on p. 41/

Polson, A. G., Bass, B. L., and Casey, J. L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature*, 380(6573):454–6. /cited on p. 41/

Popenda, M., Milecki, J., and Adamiak, R. W. (2004). High salt solution structure of a left-handed RNA double helix. *Nucleic Acids Res.*, 32(13):4044–54. /cited on p. 24/

Quyen, D. V., Ha, S. C., Lowenhaupt, K., Rich, A., Kim, K. K., and Kim, Y.-G. (2007). Characterization of DNA-binding activity of Z alpha domains from poxviruses and the importance of the beta-wing regions in converting B-DNA to Z-DNA. *Nucleic Acids Res.*, 35(22):7714–20. /cited on p. 54/

Rakus, K., Ouyang, P., Boutier, M., Ronsmans, M., Reschner, A., Vancsok, C., Jazowiecka-Rakus, J., and Vanderplasschen, A. (2013). Cyprinid herpesvirus 3: an interesting virus for applied and fundamental research. *Vet. Res.*, 44(1):85. /cited on p. 48/

Ray, B. K., Dhar, S., Henry, C., Rich, A., and Ray, A. (2013). Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family. *Cancer Res.*, 73(2):736–44. /cited on p. 37/

Ray, B. K., Dhar, S., Shakya, A., and Ray, A. (2011). Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, 108(1):103–8. /cited on p. 37/

Rebagliati, M. R. and Melton, D. A. (1987). Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. *Cell*, 48(4):599–605. /cited on p. 40/

Reineke, L. C., Kedersha, N., Langereis, M. A., van Kuppeveld, F. J. M., and Lloyd, R. E. (2015). Stress granules regulate double-stranded RNA-dependent protein kinase activation through a complex containing G3BP1 and Caprin1. *MBio*, 6(2):e02486. /cited on p. 13/

Rice, G., Patrick, T., Parmar, R., Taylor, C. F., Aeby, A., Aicardi, J., Artuch, R., Montalto, S. A., Bacino, C. A., Barroso, B., Baxter, P., Benko, W. S., Bergmann, C., Bertini, E., Biancheri, R., Blair, E. M., Blau, N., Bonthron, D. T., Briggs, T., Brueton, L. A., Brunner, H. G., Burke, C. J., Carr, I. M., Carvalho, D. R., Chandler, K. E., Christen, H.-J., Corry, P. C., Cowan, F. M., Cox, H., D'Arrigo, S., Dean, J., De Laet, C., De Praeter, C., Dery, C., Ferrie, C. D., Flintoff, K., Frints, S. G. M., Garcia-Cazorla, A., Gener, B., Goizet, C., Goutieres, F., Green, A. J., Guet, A., Hamel, B. C. J., Hayward, B. E., Heiberg, A., Hennekam, R. C., Husson, M., Jackson, A. P., Jayatunga, R., Jiang, Y.-H., Kant, S. G., Kao, A., King, M. D., Kingston, H. M., Klepper, J., van der Knaap, M. S., Kornberg, A. J., Kotzot, D., Kratzer, W., Lacombe, D., Lagae, L., Landrieu, P. G., Lanzi, G., Leitch, A., Lim, M. J., Livingston, J. H., Lourenco, C. M., Lyall, E. G. H., Lynch, S. A., Lyons, M. J., Marom, D., McClure, J. P., McWilliam, R., Melancon, S. B., Mewasingh, L. D., Moutard, M.-L., Nischal, K. K., Ostergaard, J. R., Prendiville, J., Rasmussen, M., Rogers, R. C., Roland, D., Rosser, E. M., Rostasy, K., Roubertie, A., Sanchis, A., Schiffmann, R., Scholl-Burgi, S., Seal, S., Shalev, S. A., Corcoles, C. S., Sinha, G. P., Soler, D., Spiegel, R., Stephenson, J. B. P., Tacke, U., Tan, T. Y., Till, M., Tolmie, J. L., Tomlin, P., Vagnarelli, F., Valente, E. M., Van Coster, R. N. A., Van der Aa, N., Vanderver, A., Vles, J. S. H., Voit, T., Wassmer, E., Weschke, B., Whiteford, M. L., Willemsen, M. A. A., Zankl, A., Zuberi, S. M., Orcesi, S., Fazzi, E., Lebon, P., and Crow, Y. J. (2007). Clinical and molecular phenotype of Aicardi-Goutieres syndrome. *Am. J. Hum. Genet.*, 81(4):713–25. /cited on p. 14/

Rice, G. I., Kasher, P. R., Forte, G. M. A., Mannion, N. M., Greenwood, S. M., Szynkiewicz, M., Dickerson, J. E., Bhaskar, S. S., Zampini, M., Briggs, T. A., Jenkinson, E. M., Bacino, C. A., Battini, R., Bertini, E., Brogan, P. A., Brueton, L. A., Carpanelli, M., De Laet, C., de Lonlay, P., del Toro, M., Desguerre, I., Fazzi, E., Garcia-Cazorla, A., Heiberg, A., Kawaguchi, M., Kumar, R., Lin, J.-P. S.-M., Lourenco, C. M., Male, A. M., Marques, W., Mignot, C., Olivieri, I., Orcesi, S., Prabhakar, P., Rasmussen, M., Robinson, R. A., Rozenberg, F., Schmidt, J. L., Steindl, K., Tan, T. Y., van der Merwe, W. G., Vanderver, A., Vassallo, G., Wakeling, E. L., Wassmer, E., Whittaker, E., Livingston, J. H., Lebon, P., Suzuki, T., McLaughlin, P. J., Keegan, L. P., O'Connell, M. A., Lovell, S. C., and Crow, Y. J. (2012). Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.*, 44(11):1243–8. /cited on p. 14, 44/

Rich, A., Nordheim, A., and Wang, A. H. (1984). The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.*, 53:791–846. /cited on p. 20/

Rohner, K. J., Hobi, R., and Kuenzle, C. C. (1990). Z-DNA-binding proteins. Identification critically depends on the proper choice of ligands. *J. Biol. Chem.*, 265(31):19112–19115. /cited on p. 38/

Rothenburg, S., Deigendesch, N., Dey, M., Dever, T. E., and Tazi, L. (2008). Double-stranded RNA-activated protein kinase PKR of fishes and amphibians: varying the number of double-stranded RNA binding domains and lineage-specific duplications. *BMC Biol.*, 6:12. /cited on p. 48/

Rothenburg, S., Deigendesch, N., Dittmar, K., Koch-Nolte, F., Haag, F., Lowenhaupt, K., and Rich, A. (2005). A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl. Acad. Sci. U. S. A.*, 102(5):1602–7. /cited on p. 48/

Rothenburg, S., Koch-Nolte, F., and Haag, F. (2001). DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles. *Immunol. Rev.*, 184(1):286–298. /cited on p. 37/

Saenger, W. and Heinemann, U. (1989). Raison d'être and structural model for the B-Z transition of poly d(G-C).poly d(G-C). *FEBS Lett.*, 257(2):223–7. /cited on p. 26/

Scadden, A. D. J. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat. Struct. Mol. Biol.*, 12(6):489–96. /cited on p. 16/

Schade, M., Turner, C. J., Kühne, R., Schmieder, P., Lowenhaupt, K., Herbert, A., Rich, A., and Oschkinat, H. (1999). The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 96(22):12465–70. /cited on p. 54/

Schroth, G. P., Chou, P. J., and Ho, P. S. (1992). Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem.*, 267(17):11846–55. /cited on p. 36/

Schwartz, T. (1999). Crystal Structure of the Z Domain of the Human Editing Enzyme ADAR1 Bound to Left-Handed Z-DNA. *Science (80-.)*., 284(5421):1841–1845. /cited on p. 50, 51/

Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U., and Rich, A. (2001). Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.*, 8(9):761–5. /cited on p. 53/

Seo, Y.-J., Ahn, H.-C., Lee, E.-H., Bang, J., Kang, Y.-M., Kim, H.-E., Lee, Y.-M., Kim, K., Choi, B.-S., and Lee, J.-H. (2010). Sequence discrimination of the $Z\alpha$ domain of human ADAR1 during B-Z transition of DNA duplexes. *FEBS Lett.*, 584(20):4344–50. /cited on p. 55/

Shors, T., Kibler, K. V., Perkins, K. B., Seidler-Wulff, R., Banaszak, M. P., and Jacobs, B. L. (1997). Complementation of vaccinia virus deleted of the E3L gene by mutants of E3L. *Virology*, 239(2):269–76. /cited on p. 47/

Silverman, R. H. (2007). Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response. *J. Virol.*, 81(23):12720–9. /cited on p. 47/

Simpson-Holley, M., Kedersha, N., Dower, K., Rubins, K. H., Anderson, P., Hensley, L. E., and Connor, J. H. (2011). Formation of antiviral cytoplasmic granules during orthopoxvirus infection. *J. Virol.*, 85(4):1581–93. /cited on p. 12/

Singleton, C. K., Kilpatrick, M. W., and Wells, R. D. (1984). S1 nuclease recognizes DNA conformational junctions between left-handed helical (dT-dG n. dC-dA)n and contiguous right-handed sequences. *J. Biol. Chem.*, 259(3):1963–7. /cited on p. 28/

Singleton, C. K., Klysik, J., Stirdivant, S. M., and Wells, R. D. (1982). Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature*, 299(5881):312–6. /cited on p. 28, 31/

Singleton, C. K., Klysik, J., and Wells, R. D. (1983). Conformational flexibility of junctions between contiguous B- and Z-DNAs in supercoiled plasmids. *Proc. Natl. Acad. Sci. U. S. A.*, 80(9):2447–51. /cited on p. 28/

Sommer, B., Köhler, M., Sprengel, R., and Seeburg, P. H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*, 67(1):11–19. /cited on p. 41/

Stetson, D. B., Ko, J. S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell*, 134(4):587–98. /cited on p. 14/

Stetson, D. B. and Medzhitov, R. (2006). Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity*, 24(1):93–103. /cited on p. 6/

Su, M., Han, D., Boyd-Kirkup, J., Yu, X., and Han, J.-D. J. (2014). Evolution of Alu elements toward enhancers. *Cell Rep.*, 7(2):376–85. /cited on p. 29/

Taghavi, N. and Samuel, C. E. (2013). RNA-dependent protein kinase PKR and the Z-DNA binding orthologue PKZ differ in their capacity to mediate initiation factor eIF2α-dependent inhibition of protein synthesis and virus-induced stress granule formation. *Virology*, 443(1):48–58. /cited on p. 7, 48/

Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K., Ohba, Y., and Taniguchi, T. (2007). DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature*, 448(7152):501–5. /cited on p. 45/

Tang, N., Muller, J. G., Burrows, C. J., and Rokita, S. E. (1999). Nickel and Cobalt Reagents Promote Selective Oxidation of Z-DNA. *Biochemistry*, 38(50):16648–16654. /cited on p. 32/

Taylor, D. R., Puig, M., Darnell, M. E. R., Mihalik, K., and Feinstone, S. M. (2005). New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. *J. Virol.*, 79(10):6291–8. /cited on p. 43/

Thamann, T. J., Lord, R. C., Wang, A. H., and Rich, A. (1981). The high salt form of poly(dG-dC)·poly(dG-dC) is left-handed Z-DNA: Raman spectra of crystals and solutions. *Nucleic Acids Res.*, 9(20):5443–5458. /cited on p. 21/

Thomae, R., Beck, S., and Pohl, F. M. (1983). Isolation of Z-DNA-containing plasmids. *Proc. Natl. Acad. Sci. U. S. A.*, 80(18):5550–3. /cited on p. 31/

Thomas, T. J. and Bloomfield, V. A. (1983). Chain flexibility and hydrodynamics of the B and Z forms of poly(dG-dC).poly(dG-dC). *Nucleic Acids Res.*, 11(6):1919–30. /cited on p. 34/

Trulson, M. O., Cruz, P., Puglisi, J. D., Tinoco, I., and Mathies, R. A. (1987). Raman spectroscopic study of left-handed Z-RNA. *Biochemistry*, 26(26):8624–8630. /cited on p. 24/

Tsai, W.-C. and Lloyd, R. E. (2014). Cytoplasmic RNA Granules and Viral Infection. *Annu. Rev. Virol.*, 1(1):147–70. /cited on p. 12/

Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. r. B., Sharma, S., Sirois, C. M., Jin, T., Latz, E., Xiao, T. S., Fitzgerald, K. A., Paludan, S. r. R., and Bowie, A. G. (2010). IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.*, 11(11):997–1004. /cited on p. 7/

Upton, J. W., Kaiser, W. J., and Mocarski, E. S. (2012). DAI/ZBP1/DLM-1 complexes with RIP3 to mediate virus-induced programmed necrosis that is targeted by murine cytomegalovirus vIRA. *Cell Host Microbe*, 11(3):290–7. /cited on p. 45/

Vitali, P. and Scadden, A. D. J. (2010). Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. *Nat. Struct. Mol. Biol.*, 17(99):1043–1050. /cited on p. 16/

Wahls, W. P., Wallace, L. J., and Moore, P. D. (1990). The Z-DNA motif d(TG)30 promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol. Cell. Biol.*, 10(2):785–93. /cited on p. 31/

Walker, G. T. and Aboul-ela, F. (1988). B-Z cooperativity and kinetics of poly(dG-m5dC) are controlled by an unfavorable B-Z interface energy. *J. Biomol. Struct. Dyn.*, 5(6):1209–19. /cited on p. 25/

Wang, A., Quigley, G., Kolpak, F., van der Marel, G., van Boom, J., and Rich, A. (1981). Left-handed double helical DNA: variations in the backbone conformation. *Science (80-.).*, 211(4478):171–176. /cited on p. 21/

Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, 282(5740):680–6. /cited on p. 18, 25, 32/

Wang, G., Christensen, L. A., and Vasquez, K. M. (2006). Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.*, 103(8):2677–82. /cited on p. 27, 32/

Wang, Q., Miyakoda, M., Yang, W., Khillan, J., Stachura, D. L., Weiss, M. J., and Nishikura, K. (2004). Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *J. Biol. Chem.*, 279(6):4952–61. /cited on p. 43/

Wang, Z., Choi, M. K., Ban, T., Yanai, H., Negishi, H., Lu, Y., Tamura, T., Takaoka, A., Nishikura, K., and Taniguchi, T. (2008). Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 105(14):5477–82. /cited on p. 45, 46/

Ward, S. V., George, C. X., Samuel, C. E., and Oldstone, M. B. A. (2011). Reply to Steinman and Wang: Involvement of the p150 isoform of the RNA editing enzyme ADAR1 in embryonic lethality. *Proc. Natl. Acad. Sci. U. S. A.*, 108(24):E200–E200. /cited on p. 43/

Watson, J. C., Chang, H. W., and Jacobs, B. L. (1991). Characterization of a vaccinia virus-encoded double-stranded RNA-binding protein that may be involved in inhibition of the double-stranded RNA-dependent protein kinase. *Virology*, 185(1):206–16. /cited on p. 47/

Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–8. /cited on p. 18/

Whitaker-Dowling, P. and Youngner, J. S. (1984). Characterization of a specific kinase inhibitory factor produced by vaccinia virus which inhibits the interferon-induced protein kinase. *Virology*, 137(1):171–181. /cited on p. 47/

White, J. P., Cardenas, A. M., Marissen, W. E., and Lloyd, R. E. (2007). Inhibition of cytoplasmic mRNA stress granule formation by a viral proteinase. *Cell Host Microbe*, 2(5):295–305. /cited on p. 12/

Wilkins, M. and Randall, J. (1953). Crystallinity in sperm heads: Molecular structure of nucleoprotein in vivo. *Biochim. Biophys. Acta*, 10:192–193. /cited on p. 18/

Willis, K. L., Langland, J. O., and Shisler, J. L. (2011). Viral double-stranded RNAs from vaccinia virus early or intermediate gene transcripts possess PKR activating function, resulting in NF-kappaB activation, when the K1 protein is absent or mutated. *J. Biol. Chem.*, 286(10):7765–78. /cited on p. 47/

Wölfl, S., Wittig, B., and Rich, A. (1995). Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim. Biophys. Acta - Gene Struct. Expr.*, 1264(3):294–302. /cited on p. 31/

Wong, B., Chen, S., Kwon, J.-A., and Rich, A. (2007). Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces

cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.*, 104(7):2229–34. /cited on p. 33, 35/

Wu, C., Hu, Y., Fan, L., Wang, H., Sun, Z., Deng, S., Liu, Y., and Hu, C. (2015). Ctenopharyngodon idella PKZ facilitates cell apoptosis through phosphorylating eIF2α. *Mol. Immunol.*, 69:13–23. /cited on p. 48/

Wu, J. and Chen, Z. J. (2014). Innate immune sensing and signaling of cytosolic nucleic acids. *Annu. Rev. Immunol.*, 32:461–88. /cited on p. 4/

Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., and Chen, Z. J. (2013). Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science (80-.).*, 339(6121):826–30. /cited on p. 6/

Wu, T. T., Bichko, V. V., Ryu, W. S., Lemon, S. M., and Taylor, J. M. (1995). Hepatitis delta virus mutant: effect on RNA editing. *J. Virol.*, 69(11):7226–31. /cited on p. 43/

Xia, P., Wang, S., Ye, B., Du, Y., Huang, G., Zhu, P., and Fan, Z. (2015). Sox2 functions as a sequence-specific DNA sensor in neutrophils to initiate innate immunity against microbial infection. *Nat. Immunol.*, 16(4):366–75. /cited on p. 15/

Yang, X. L. and Wang, A. H. (1997). Structural analysis of Z-Z DNA junctions with A:A and T:T mismatched base pairs by NMR. *Biochemistry*, 36(14):4258–67. /cited on p. 29/

Yang, Y.-G., Lindahl, T., and Barnes, D. E. (2007). Trex1 Exonuclease Degrades ssDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease. *Cell*, 131(5):873–886. /cited on p. 14/

Yuwen, H., Cox, J. H., Yewdell, J. W., Bennink, J. R., and Moss, B. (1993). Nuclear localization of a double-stranded RNA-binding protein encoded by the vaccinia virus E3L gene. *Virology*, 195(2):732–44. /cited on p. 47/

Zaborske, J. M., Narasimhan, J., Jiang, L., Wek, S. A., Dittmar, K. A., Freimoser, F., Pan, T., and Wek, R. C. (2009). Genome-wide analysis of tRNA charging and activation of the eIF2 kinase Gcn2p. *J. Biol. Chem.*, 284(37):25254–67. /cited on p. 10/

Zacharias, W., Jaworski, A., and Wells, R. D. (1990). Cytosine methylation enhances Z-DNA formation in vivo. *J. Bacteriol.*, 172(6):3278–83. /cited on p. 23/

Zang, Z. and Olson, W. (1987). A Model for the B→Z Transition of DNA Involving Solitary Excitations. In Bishop, A., Campbell, D., Kumar, P., and Trullinger, S., editors, *Nonlinearity Condens. Matter SE - 30*, volume 69 of *Springer Series in Solid-State Sciences*, pages 265–270. Springer Berlin Heidelberg. /cited on p. 25/

Zarling, D. A., Calhoun, C. J., Hardin, C. C., and Zarling, A. H. (1987). Cytoplasmic Z-RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 84(17):6117–21. /cited on p. 37/

Zhang, S., Lockshin, C., Herbert, A., Winter, E., and Rich, A. (1992). Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae. *EMBO J.*, 11(10):3787–96. /cited on p. 38/

2

ORF112 a novel member of Zα containing protein family

Abstract

Z-DNA/Z-RNA binding domains (Zαs) possess the ability to recognize and interact with nucleic acids in the left-handed conformation. Zα domains are components of proteins involved in the antiviral pathway known as interferon response. In addition, the Zα domain of the poxviral protein E3L is required for the full pathogenesis of vaccinia virus *in vivo*. Here, we provide evidence that another viral family of cyprinid herpesviruses encodes also a protein (ORF112) with a functional Zα domain. Our biochemical assays revealed that Zα ORF112 indeed interacts with Z-DNA. Moreover, the structure (at 1.75 Å) of free Zα ORF112 uncovers a fold similar to other members of Z-DNA/Z-RNA binding domain family. Surprisingly, free Zα ORF112 dimerizes through a domain-swapping mechanism. Overall, this work demonstrates that Zα ORF112 is a novel member of Z-DNA/Z-RNA binding family and suggested that it may be involved in the subversion of the antiviral responses of the infected host, similarly to the poxviral E3L.

Publication

Ana Rita Tomé[§], Krzysztof Kuś[§], Silvia Correia, Lara Martins Paulo, Sónia Zacarias, Matteo de Rosa, Delio Figueiredo, R. Michael E. Parkhouse and Alekos Athanasiadis, Crystal Structure of a Poxvirus-Like Zalpha Domain from Cyprinid Herpesvirus 3, *Journal of Virology*, **2013**, 87(7), p.3998-4004

[§] Both authors contributed equally to this work

Contribution

Alekos Athanasiadis initiated this study. Ana Rita Tomé and Krzysztof Kuś contributed equally to this work. Ana Rita Tomé purified and crystallized Zα ORF112. Krzysztof Kuś solved the structure of Zα ORF112 (with assistance from Alekos Athanasiadis) and performed band shift assays (excluding competition assay).

2.1 Introduction

Three major conformations have been attributed to nucleic acids helices: A, B and Z. Z-RNA and Z-DNA are the left-handed conformers and were named after a characteristic zig-zag trace of the phosphate/sugar nucleic acid backbone (Wang *et al.*, 1979). The class of the winged helix-turn-helix domains termed Z-DNA/Z-RNA binding domains (Zαs) has the unique ability to interact with this left-handed conformation of either DNA or RNA (reviewed (Athanasiadis, 2012)). A typical Zα domain is described by a $\alpha1\beta1\alpha2\alpha3\beta2\beta3$ topology. Both $\alpha3$ helix and the wing region (the loop between $\beta2$ and $\beta3$ strands) contribute to the interaction with the nucleic acids. The Zα domain recognizes the shape of Z-conformation but it is still under debate whether it can induce the B-to-Z transition or selectively interact with the left-handed form. Despite relatively low sequence identity Zα domains are very similar on the structural level. The structure of the first Zα from ADAR1 revealed a set of residues critical for the interaction and recognition of the left-handed conformation. It was shown that tyrosine (Y177), asparagine (N173) and tryptophan (W195) are required for the recognition of Z-DNA/Z-RNA (**Figure 1.15**)(Schwartz, 1999; Placido *et al.*, 2007). Mutations in any of these residues abolish binding of the Z-conformation. Currently, Zα domains are found in proteins which are either involved in the host interferon response pathway (ADAR1, DAI and PKZ) or are a part of the viral evasion mechanism (poxviral E3L). DAI and PKZ are implicated in the cytoplasmic recognition of the foreign nucleic acids and initiation of

the signaling cascades that result in the production of interferons. On the other hand, E3L is a viral protein with Zα and dsRBD domains, and both are required for the full pathogenesis of poxviruses in a mouse infection model. Interestingly, the substitution of Zα E3L by other Z-DNA/Z-RNA binding domain (from other proteins) restores the viral pathogenicity. Until now, poxviruses were the only known viral family with a Zα domain containing protein (Kim *et al.*, 2003).

Intriguingly, the teleost fish species of the Cypriniformes (common carp, koi and goldfish) and Salmoniformes (i.e. Atlantic salmon) possess a gene encoding a protein paralogous to PKR named PKZ (Rothenburg *et al.*, 2005). In contrast to PKR, PKZ instead of dsRBD domains has two Z-DNA/Z-RNA binding domains. In the case of PKR, dsRBD domains detect cytoplasmic dsRNA, whereas PKZ is postulated to sense DNA or RNA with a propensity to form the left-handed conformation. Indeed, *in vitro* studies have demonstrated that PKZ phosphorylates eIF2α efficiently in the presence of Z-forming sequence but not poly(I:C) RNA (Bergan *et al.*, 2008). Detection of the nucleic acid by PKZ or PKR would lead to conformational changes that render these proteins active to phosphorylate eIF2α which in turn shuts down the translation. In fish, PKZ and PKR seem to act cooperatively in anti-viral interferon responses (Liu *et al.*, 2011).

Cyprinid herpesvirus 3 (CyHV3) belongs to the order of Herpesvirales and the recently created *Alloherpesviridae* family (Rakus *et al.*, 2013). According to the International Committee on Taxonomy of Viruses, *Alloherpesviridae* family consists of 12 viruses (10 infecting fish species and 2 infecting amphibians). CyHV3 infects a range of fish species but only common and koi carp develop disease symptoms. As common carp is one of the most valuable agricultural fish species, CyHV3 is responsible for the severe economic losses and has a heavy ecological impact on natural populations. Interestingly, other species for which CyHV3 infection is asymptomatic can transmit the disease to the ones manifesting disease symptoms (Perelberg *et al.*, 2003; Rakus *et al.*, 2013). CyHV3 has an icosahedral capsid which protects a linear, single copy,

double-stranded DNA genome. It possesses the largest genome (295 kbp) found in herpesviruses (Michel *et al.*, 2010). The capsid is surrounded by a protein-rich layer, named tegument. Additionally, the virus is coated with an envelope derived from the host membranes (Miwa *et al.*, 2007). Surprisingly, studies of the genomes of cyprinid herpesvirus 3 uncovered at least four genes encoding proteins with the signature of poxviruses (Ilouze *et al.*, 2006). Thus, it was proposed that poxviruses and alloherpesviruses could have a common ancestor or that these genes were acquired through horizontal gene transfer. The genome of CyHV3 was sequenced which shed light on the repertoire of the genes potentially involved in the replication and manipulation of the anti-viral responses. We detected a cyprinid herpesviral protein with the signature of Zα domain in the ORF112 gene product. As until now Zα domains were not found in any other viral family than poxviruses, we decided to characterize this domain as this could uncover a novel immune evasion protein for these viruses. In this chapter, we will provide a biochemical and structural data demonstrating that Zα ORF112 is a functional Z-DNA/Z-RNA binding domain. It is a novel Zα domain which shares similarities with Zα E3L and Zα PKZ. Our study suggests that poxviruses and *Alloherpesviridae* may utilize related mechanisms to counteract the host interferon responses.

2.2 Materials and Methods

Cloning and expression

Sequences encompassing M221-A278, S219-A278, and N216-A278 of ORF112 (BAF48926.1) were amplified from CyHV-3 genomic DNA (kindly provided by the Friedrich-Loeffler Institute) and cloned into a pET28a vector as a fusion with a His-tag (six histidines) at the NheI-XhoI restriction sites. Protein expression was induced in *Escherichia coli* BL21(DE3) cells with the addition of 0.4 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 3 h. Cell pellets from 1-liter cultures were lysed with BugBuster (Novagen) in the presence of 1 mM phenylmethylsulfonyl

fluoride (PMSF) and Benzonase (Novagen) for 2 h at 4 °C. The proteins were loaded onto a HiTrap nickel-nitrilotriacetic acid (Ni-NTA) column on an Äkta purifier system and eluted using a gradient between 20 mM and 300 mM imidazole. The first construct did not result in any detectable protein production. The eluted protein from the other two constructs was dialyzed against 0.5x thrombin buffer (10 mM Tris [pH 8.4], 75 mM NaCl, 1.25 mM CaCl2) and incubated with thrombin overnight at 4 °C. The cleaved proteins were directly applied to a MonoS column and then eluted with a 75-to-600 mM NaCl gradient. Selected fractions were concentrated and buffer exchanged to a final concentration of 7.5 mg/ml in 20 mM Tris (pH 7.5), 50 mM KCl or 10 mM HEPES, 50 mM NaCl (pH 7.4) for use in crystallization and biochemical characterization.

Protein crystallization

The purified protein was used for crystallization screens (Crystal Screen I-II from Hampton Research and the Joint Center for Structural Genomics from Molecular Dimensions) in complex with $T(CG)_3$ duplex oligonucleotides, and crystals were obtained under several conditions. However, in all cases the crystals were also reproducible in controls (in the absence of nucleic acids), suggesting that only the protein component had been crystallized. While we continued our efforts to obtain crystals of the ORF112/DNA complex we decided to characterize protein-only crystals as this would provide valuable knowledge confirming or not the $Z\alpha$-like structure of ORF112. Highest-quality rod-shaped crystals (around 0.2 by 0.05 by 0.05 µm) were obtained using a reservoir containing 0.2 M NaCl, 0.1 M cacodylate (pH 6.5), and 2 M ammonium sulfate. Well-shaped crystals were subsequently frozen at 100 K in the presence of Paratone N or 20% glycerol as a cryoprotectant.

Data collection and structure determination

Flash-frozen crystals at 100 K were exposed to X-rays at the ID23-1 beamline of ESRF/Grenoble using a 0.97 Å wavelength. The best crystal diffracted up to 1.76 Å. Data processing was performed using the XDS software package (Kabsch, 2010). The ORF112 $Z\alpha$ crystals belong to the

orthorhombic $P2_12_12_1$ space group with the following parameters: a = 51.7 Å, b = 54.6 Å, c = 86.6 Å, and $\alpha = \beta = \gamma = 90.00°$. Phasing was performed using molecular replacement with ADAR1 Zα (PDB code 1QBJ) as the starting model was successful only when a C-terminally truncated model was used, leading to a solution with four monomers per asymmetric unit. The structure was refined from the starting model using Phenix (Adams *et al.*, 2010) and iterative rebuilding using Coot (Emsley *et al.*, 2010). The resulting electron density clearly showed the C-terminus of the protein leading to a neighboring monomer, suggesting domain-swapping of the last β-strand (with 10 amino acids being exchanged) and resulting in an arrangement of two dimers per asymmetric unit. The final refined model has an R_{work}/R_{free} ratio of 20.1/24.7 and comprises 248 protein residues and 213 solvent molecules, among which nine are sulfate ions (see **Table A.1** for statistics included in **Appendix**). The backbone of P217-A278, S215-P277, H212-A278, and S219-A278 from chains A, B, C, and D, respectively, was visible in the electron density map. Some side chains of solvent-exposed amino acids (S215B, N216B, H212C, M213C, R266C, R266D, and Q270D) had no visible electron density, while the side chain of H188C shows two alternative conformations which were both modeled. The nine sulfate ions could be located in the structure in similar positions in each the four monomers. The Zα domain crystal structure determined here has been deposited in the RCSB database (RCSB code RCSB075715; PDB code 4HOB).

DNA binding assays

The ability of the purified protein to bind DNA was evaluated by gel mobility shift analysis using a $T(CG)_3$ duplex oligonucleotide (Integrated DNA Technologies). Mixtures of protein with 50 µM DNA at ratios of 1/1 to 4/1 were incubated at 37 °C for 20 min and then subjected to electrophoresis on non-denaturing polyacrylamide gels and stained first with SYBR followed by Coomassie blue. A distinct band corresponding to the protein-DNA complex was clearly visible. Complete conversion to Z-DNA could be observed at protein/DNA ratios of 2:1. The conformation of the oligonucleotides in the complex was evaluated using circular

dichroism (CD) spectroscopy on a Jasco J-815 CD system in a 0.1-mm cuvette. A clear inversion of the trace, characteristic of the Z-DNA helix, was observed at a wavelength of 255 nm (see **Figure 2.2B**).

Size exclusion chromatography characterization

We used size exclusion chromatography to characterize the oligomerization state of the protein in solution. An S75 prepacked column (GE Healthcare) was equilibrated with 50 mM Tris (pH 7.5) and 100 mM KCl and calibrated using protein standards (molecular mass, 6,500 to 66,000 Da; Sigma-Aldrich). ORF112 Zα protein (1 mg) in a volume of 200 µl was loaded on the column and the elution profile was recorded. Calibration of the column was performed with standards (Sigma). A similar procedure was carried out for the protein in the buffer containing ammonium sulfate (50 mM).

2.3 Results

2.3.1 ORF112 as a putative Zα containing protein from cyprinid herpesviruses

Fast developments in the DNA sequencing methodologies result in a growing number of sequenced genomes. Viral genomes of cyprinid herpesviruses were sequenced and provide valuable information to study their pathogenicity. Using homology-based methods to find novel members of the Zα domain family, we discovered a gene named *ORF112* in the cyprinid herpesvirus 3 genome (BAF48926.1) that contained a segment with the Zα signature. This protein is predicted to have 278 amino acids and a potential Zα domain is located in the C-terminal part. As ORF112 could be a potential homolog of the poxviral E3L, we inspected whether the N-terminus of the protein has a signature of the dsRBD domain. We could not detect any recognizable motif. In fact, the N-terminal part of ORF112 contains stretches of low complexity regions with high prevalence of glutamine. We hypothesized that a shorter version of the protein starting from the internal methionine (92 amino acids, mostly covering Zα domain) would be a functional form (**Figure 2.1A**). However, our more recent findings advocate against this idea as mass spectrometry analysis of the CyHV3 proteome indicated a presence of peptides matching N-terminal part (compare **Chapter 3**). To gain insight into the evolutionary link between Zα domains we compared the sequence of Zα ORF112 with other members of Zα family. We found that the highest identity is shared between Zα ORF112 and Zα domains from PKZ (*Gobiocypris rarus*, *Danio rerio*) indicating a potential functional and evolutionary link between these domains from PKZ and ORF112. Based on the sequence analysis, we noted that all critical residues involved in the binding of the Z-DNA/Z-RNA are conserved. Therefore, residues Tyr257, Asn253, Trp274 of ORF112 are equivalent to Tyr177, Asn173 and Trp195 of ADAR1. In addition, residues that are part of the hydrophobic core important for the proper folding of the domain are preserved in Zα

ORF112 (**Figure 2.1B**). Thus, all our preliminary analysis suggested that ORF112 contains a genuine Zα domain.

(A)

ORF112:

MASDTNGTGT SNGTYLTLRK KKSEPKSVTF NLADDGARVV TARGTSSAAL ASSLLSSENA RLQLQIQQRQ

IQQQLQQLEL REFQDLKAQT PLQQPQPQQH QCQCQCHHHK HDNRGHGHSH GHGHRGGKRG QLKRQNASNN

LLAQPFRPPK EPSTEVVVSQ PQMKPPPLPL QSLREPLDVD AVLIPRMSVD DLDGLDGAEK VKTTASEAIP

ALPRLNPISE EMNLKILAYL GTKQGAKAVH IAQSLGAQRS EVNRHLYRMS EDGRVRKHPQ HPVWYLPA

(B)

Figure 2.1: Identification of ORF112 from cyprinid herpesvirus 3 as a putative Zα domain-containing protein. (A) ORF112 is predicted to be 278 amino acids long protein with a repetitive N-terminal part (accession number: BAF48926.1). An alternative product starting at residue 187 (methionine in yellow) is highlighted by a gray shading. Recent mass spectrometry results show the presence of the peptides from N-terminal part suggesting that the full-length protein is indeed produced. The starting residues of the Zα constructs used in this study are highlighted (green shading). **(B)** Alignment of Zα ORF112 CyHV3 with PKZ, E3L and ADAR1 Zα domains. Abbreviation preceding names of proteins: h, *Homo sapiens*; vv, vaccinia virus; ss, *Salmo salar*; ca, *Carassius auratus*; ci, *Ctenopharyngodon idella*; dr, *Danio rerio*; gr, *Gobiocypris rarus*. The of critical residues involved in recognition of Z-DNA/Z-RNA are marked with arrows and the numbering refers to the residues of ORF112 (corresponding to Asn174, Tyr177, and Trp195 in hADAR1). The fully conserved residues are presented in red shading and blue boxes indicate conservative changes. The dark squares under the alignment mark residues forming the hydrophobic core of the protein.

2.3.2 Biochemical characterization of Zα ORF112

First, we attempted to express the full-length ORF112 in the bacterial expression system but the protein was present almost exclusively in the insoluble fraction (data not shown). Two His-tag constructs (regions: S219-A278 and N216-A216) of ORF112 yielded soluble proteins but the shorter version was temperature-sensitive and incubation at 37 °C induced precipitation of the protein. Therefore, the longer more stable form was used for biochemical assays and crystallization trials. To evaluate whether Zα ORF112 can interact with the left-handed DNA we used electrophoretic mobility shift assay (EMSA) with the minimal Zα substrate T$(CG)_3$. **Figure 2.2A** demonstrates that purified Zα ORF112 interacts with the CpG repeat and the induced band shift is comparable to that of Zα ADAR1 when interacting with the same oligonucleotide. On the other hand, when the assay is performed with the oligonucleotide (TA2) with a low propensity to form Z-DNA we do not detect specific complexes (**Figure 2.2A**). Circular dichroism (CD) is a well-established methodology to assess the presence of the left-handed conformation (see **Chapter 1**). We performed CD experiments for different protein/DNA ratios. In the presence of Zα ORF112, we observe a typical inversion of the DNA spectrum characteristic for the Z-conformation (**Figure 2.2B**).

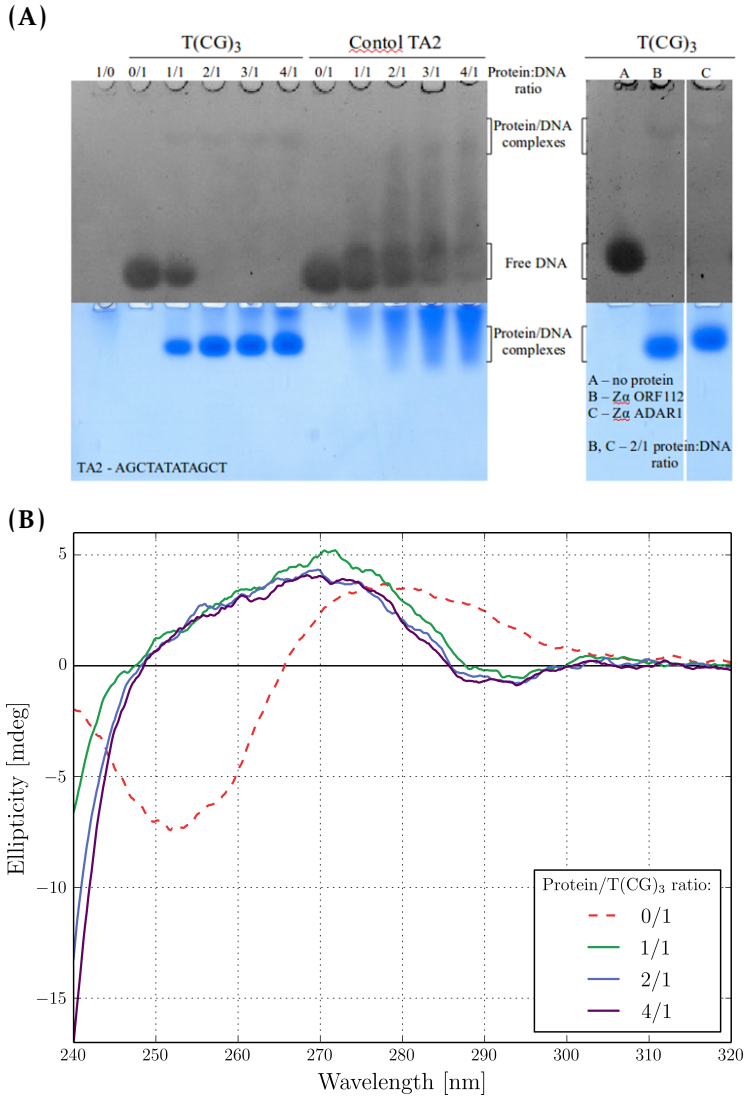


Figure 2.2: Zα ORF112 binds $T(CG)_3$ duplex repeat in the left-handed conformation. (A) An electrophoretic mobility shift assay (EMSA) demonstrates that Zα ORF112 forms a complex with a $T(CG)_3$ duplex as shown on the native polyacrylamide gel stained for the nucleic acids with RedSafe (top) and proteins with Coomassie blue (bottom). No specific bands are formed for Zα ORF112 with control TA2 duplex (low propensity to form Z-DNA). The molar ratio of protein to DNA is indicated above the lanes. The right side of the panel (A) includes a result of the EMSA assay for the prototypical Zα ADAR1 with $T(CG)_3$ (protein/DNA ratio:2/1). **(B)** Circular dichroism (CD) was measured for different Zα ORF112/$T(CG)_3$ ratios (1/1, 2/1 and 4/1). CD spectrum of the DNA duplex alone was also recorded (B-DNA control; red, dashed line). A characteristic shift of the ellipticity at 255 nm (compared to B-DNA control) indicates Z-DNA formation.

Having confirmed that Zα ORF112 has the ability to interact with the left-handed DNA, we decided to assess a potential role of Zα ORF112 as a competitive inhibitor of PKZ. We expressed and purified the first of the two Z-DNA binding domains of *Danio rerio* PKZ (drPKZ) which has been previously studied in our laboratory (de Rosa *et al.*, 2013). The first Zα drPKZ is a stronger binder than the second domain (de Rosa *et al.*, 2013). In a competition assay where equimolar amounts of both domains were simultaneously present, a complex migrating as Zα ORF112/DNA was preferentially formed (**Figure 2.3**). This suggests that under these conditions Zα ORF112 can outcompete Zα drPKZ.

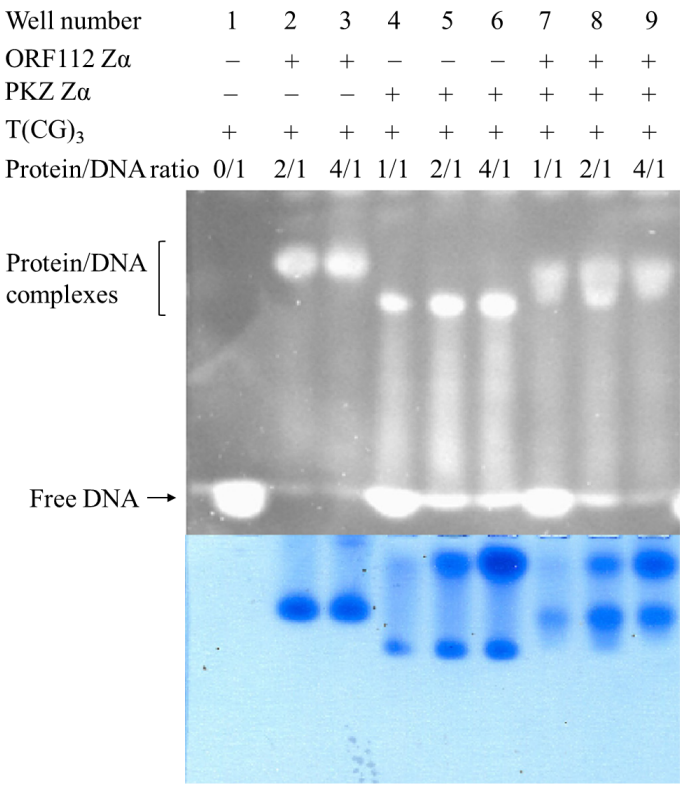


Figure 2.3: *In vitro* competition between Zα ORF112 and Zα PKZ for $T(CG)_3$. The band shift of $T(CG)_3$ in the presence of Zα ORF112 (lanes 2,3) and *Danio rerio* Zα1 PKZ (lanes 4, 5, 6). Lanes 7, 8, 9 reflect the complexes formed in the presence of equimolar amounts of both domains. Protein/DNA ratios are indicated above the gel and correspond to total protein to DNA ratio. DNA has been stained with RedSafe (top) and protein with Coomassie blue stain (bottom).

2.3.3 Structure of free Zα ORF112

While our initial crystallization trials of Zα ORF112 with $T(CG)_3$ failed to yield crystals of the complex, crystals of free protein were readily obtained and the structure of the apoprotein was solved at a resolution of 1.75 Å using a truncated Zα ADAR1 (PDB: 1QBJ) as a starting model for molecular replacement. The asymmetric unit consists of four Zα ORF112 monomers which are organized into two non-identical dimers. Differences between dimers concern the solvent pattern, the angle, the distance between the two monomers (**Figure 2.4**).

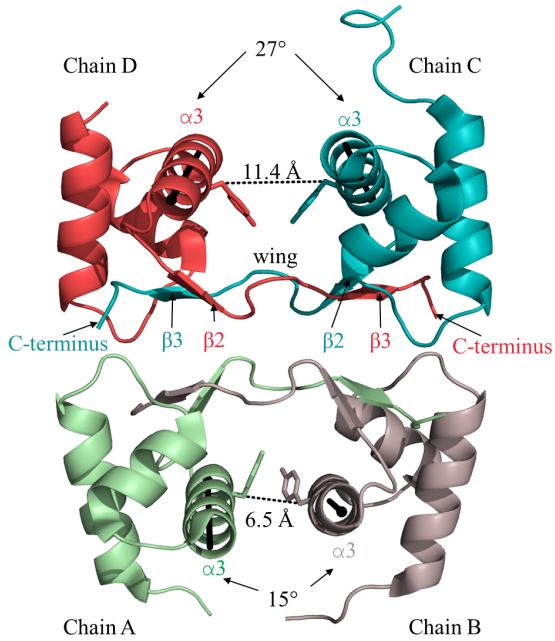


Figure 2.4: The asymmetric unit of the Zα ORF112 structure. The asymmetric unit (ASU) consists of four monomers of Zα ORF112. Each two monomers form dimers through domain-swapping. The exchange involves the two last β-strands of each domain and $\beta 2$ of one monomer interacts with $\beta 3$ of the other monomer. The two dimers present in ASU are not identical and have different angles between monomers at the hinge region, which is reflected in different distances between monomers. The distance between Cβ atoms of tyrosines 257 of each dimer is shown as a reference. Relative angles between $\alpha 3$ helices of each dimer are included (measured between black vectors depicted inside $\alpha 3$ helices).

The overall fold of each of the monomers is similar to the prototypical Zα domain of ADAR1 and belongs to the wHTH structural motif. Unexpectedly, each monomer in the dimer exchanges parts of its

C-terminus ("domain-swapping") a behavior that has never been observed for Zα domains. Each of the monomers uses the terminal amino acids (from the β2-strand, the wing region β3-strand) to engage in the reciprocal interaction with the neighboring chain. One can observe that β2-strand of one monomer is in the proximity of β3-strand of the other. This unusual dimer suggests that the wing region in the free protein has a higher degree of flexibility than it was anticipated from the previous structures.

The mechanism of domain-swapping

To understand how the ORF112 dimer could potentially bind Z-DNA we superimposed Zα ORF112 with Zα ADAR1 bound to Z-DNA. To our surprise, a sulfate ion occupied the exact position of the phosphate (from the DNA backbone) that forms the crucial for the recognition interaction with the amino acid triplet (**Figure 2.5**).

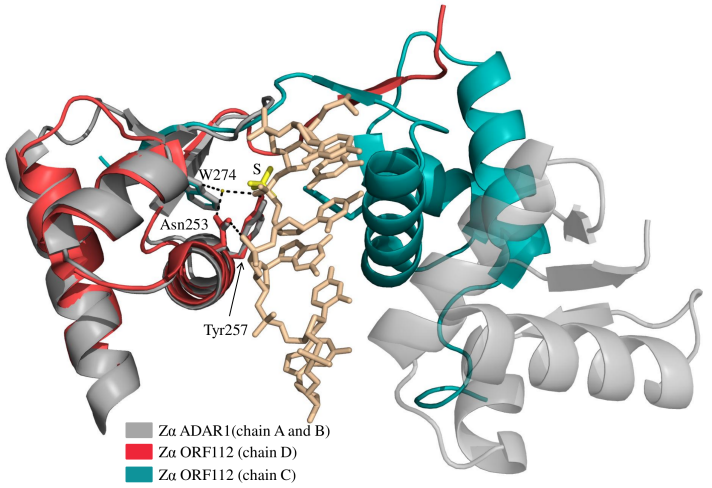


Figure 2.5: The domain-swapped Zα ORF112 dimer is not positioned to bind the Z-DNA duplex. Zα ORF112 (chain D, red ribbon) was aligned to Zα ADAR1 complexed with a T(CG)$_3$ duplex oligonucleotide (chain A, a gray ribbon, PDB:1QBJ). Additionally, the figure shows chain C of Zα ORF112 (a dark cyan ribbon) and chain B from Zα ADAR1 structure (a semi-transparent, gray ribbon). For clarity purposes, only one Z-DNA strand is included and is represented as a stick model (wheat color). The sulfate ion (a yellow stick model, marked as S) is found at the same location as a phosphate group from Z-DNA backbone which is important for the Z-DNA recognition.

The solution used for the crystallization contained 2M ammonium sulfate and we located in the electron density several sulfate ions bound to the positively charged regions of the domains. This observation indicated that sulfate ions can be important for the dimer formation. To evaluate the involvement of the sulfates in the dimer formation we performed gel filtration chromatography experiments to study oligomerization state of the protein in different buffers. In the buffer without sulfate ions (50 mM TRIS pH 7.5, 100 mM KCl) protein eluted as a single peak with a mass of 7.7 kDa (close to calculated mass of monomer; **Figure 2.6A**). A similar experiment in buffer containing 50 mM sulfate resulted in the appearance of the small second peak (14.2 kDa in agreement with dimer mass) (**Figure 2.6B**). We observed also the third peak which may reflect a higher order oligomeric state of the protein. All collected fractions were analyzed by SDS-PAGE and resolved as proteins with identical molecular mass, supporting the idea that gel filtration results correspond to different oligomeric states of Zα ORF112 (**Figure 2.6C**).

Modeling of the Z-DNA binding by the dimeric Zα ORF112 clearly demonstrated that the dimer as found in the crystal structure cannot bind the Z-DNA due to the steric hindrance (**Figure 2.5**). As we could not exclude the possibility that the flexible nature of the hinge region would allow the repositioning of the monomers forming Z-DNA binding competent dimer. To test the Z-DNA binding by the dimeric and monomeric Zα ORF112 forms we carried out EMSA assay with the two main fractions collected from gel filtration experiments (**Figure 2.6D**). It should be noted that a full separation of the peaks from gel filtration was not achievable. Nevertheless, we observed that the dimer fraction resulted in the weaker band shift compared to the monomer, indicating that this fraction does not have the full Z-DNA binding capacity (**Figure 2.6D**). As now, we have in our disposition the structure of Zα ORF112 bound to Z-DNA we will revisit the domain-swapping dimerization in discussion of **Chapter 3**.

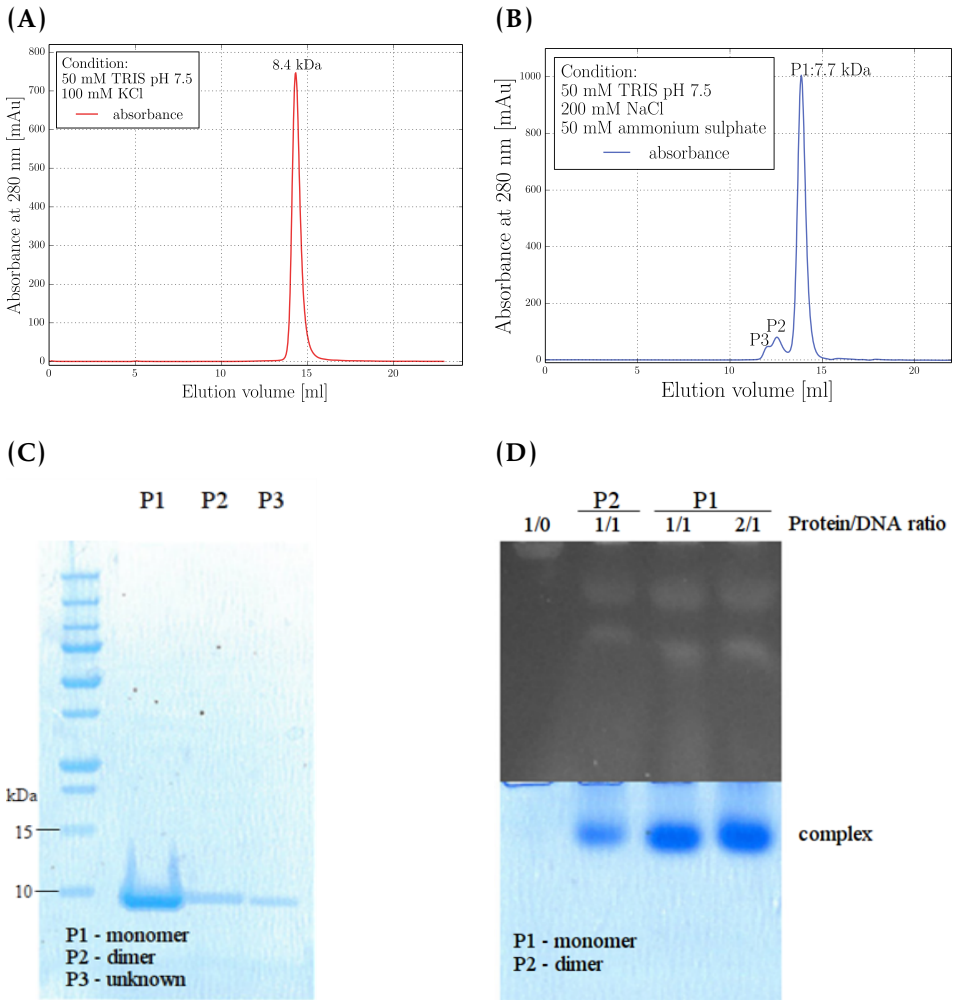


Figure 2.6: Ammonium sulfate induces Zα ORF112 dimerization.
Gel filtration (performed on a Superdex 75 column) results of Zα ORF112 in the absence **(A)** or presence **(B)** of sulfate ions. The addition of 50 mM ammonium sulfate induces a dimerization. P1 - first peak (monomer; 7.7 kDa), P2 - second peak (dimer; 14.2 kDa), P3 - third peak (unknown).
(C) Denaturing SDS-PAGE gel of samples corresponding to collected peaks shown in the panel (B). All peaks have the same molecular mass indicating that the collected fractions correspond to different oligomeric states of Zα ORF112.
(D) Gel shift mobility assay with the $T(CG)_3$ duplex oligonucleotide and Zα ORF112 from the gel filtration experiments with ammonium sulfate panel (B). The dimer form (P2) presents a weaker band shift than monomeric form (P1).

2.4 Discussion

Until this work, the only known viral protein containing a Zα domain was the interferon response inhibitor E3L from poxviruses. Our structural and biochemical results affirm that the cyprinid herpesvirus 3 ORF112 protein contains a functional Zα domain. Cyprinid herpesvirus 3 belongs to the *Alloherpesviridae* family within the order of *Herpesvirales*. Interestingly, other herpesviruses are not known to encode any protein with Zα domain signature. Therefore, the presence of Zα domains in *Alloherpesviridae* and poxviruses would suggest either a common ancestor or horizontal gene transfer between these viruses. Beyond ORF112 few other genes are found to be shared between these viral families (thymidylate monophosphate kinase, thymidine kinase, ribonucleotide reductase) emphasizing that an evolutionary link exists between them (Ilouze *et al.*, 2006). In addition, it has been confirmed that poxviruses can infect fish species (at least in the agriculture) (Gjessing *et al.*, 2015). Despite, these apparent parallels between these viral families, in the next chapter, we would argue that Zα ORF112 is structurally and phylogenetically more similar to Zα PKZ. Thus, our data suggest that if Zα ORF112 and E3L have the common ancestor the split between them happened in the distant past.

The *in vivo* experiments have shown that both domains of the poxviral E3L are required for the full pathogenesis. The dsRBD domain of E3L acts as an inhibitor of PKR and Zα E3L is thought to antagonize the Zα domains of DAI. Interestingly, fish species of the orders Cypriniformes and Salmoniformes (e.g. zebrafish, common carp and koi) possess PKR and its paralog PKZ (with two Z-DNA binding domains). It has been demonstrated that PKZ (similarly to PKR) can *in vitro* phosphorylate eIF2α in response to Z-forming sequences but not dsRNA (poly I:C used for PKR) (Yang *et al.*, 2011). Our *in vitro* competition assay between Zα ORF112 and PKZ showed that the viral Z-DNA binding domain can outcompete the host domain in the nucleic acids binding. Consequently, the mechanism by which Zα ORF112 may be evading the host cell antiviral responses may be exerted by the inhibition of PKZ activation

through masking of the viral nucleic acids. On the other hand, as we observed that the domain-swapped dimer due to the steric hindrance cannot bind double-stranded Z-DNA, we hypothesized an alternative mechanism of Zα ORF112 action which involves a formation of inactive PKZ-ORF112 heterodimers. However, our structure of Zα ORF112 and Z-DNA (described in the following chapter) would advocate that domain-swapping albeit very informative for mechanistic purposes, is a crystallization artifact.

Remarkably, all Zα domains share a similar fold, even though their sequence identity is relatively low. Zα ORF112 has a typical Z-DNA/Z-RNA binding domain fold (wHTH) with the key residues important for the recognition of left-handed nucleic acid conformation and for folding of the hydrophobic core conserved. Moreover, gel shift and circular dichroism experiments confirmed the Z-DNA binding activity of Zα ORF112. The crystal structure of the free Zα ORF112 demonstrated a dimer between two chains formed by a domain-swapping mechanism. This dimer formation is facilitated by a sulfate found in the exact same position where critical DNA phosphate is expected in the complex. These results suggest that DNA binding may also lead to conformational changes of the wing region allowing for induced fit.

In summary, we provided the first evidence that Zα ORF112 is a Z-DNA/Z-RNA binding domain and consequently that ORF112 is a candidate host-responses evasion protein. Our study extends the family of the viral proteins containing Zα domain beyond poxviruses. These findings are also interesting from the evolutionary perspective as they suggest evolutionary links between poxviruses and alloherpesviruses. The next chapter contains a further analysis of Zα ORF112 by examining its binding mechanism to Z-DNA.

Acknowledgements

I owe my deepest gratitude to my supervisor Alekos Athanasiadis for all the support and help with meanders of structural biology. I would like to thank Ana Rita Tomé for all her contributions to this work. I would like to acknowledge Alekos Athanasiadis and Ewa Chrostek for the reading of this chapter.

Bibliography

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, 66(Pt 2):213–21. /cited on p. 90/

Athanasiadis, A. (2012). Zalpha-domains: at the intersection between RAthanasiadis, A., 2012. Zalpha-domains: at the intersection between RNA editing and innate immunity. Seminars in Cell & Developmental Biology.NA editing and innate immunity. *Semin. Cell Dev. Biol.*, 23(3):257–280. /cited on p. 86/

Bergan, V., Jagus, R., Lauksund, S., Kileng, O. y., and Robertsen, B. r. (2008). The Atlantic salmon Z-DNA binding protein kinase phosphorylates translation initiation factor 2 alpha and constitutes a unique orthologue to the mammalian dsRNA-activated protein kinase R. *FEBS J.*, 275(1):184–97. /cited on p. 87/

de Rosa, M., Zacarias, S., and Athanasiadis, A. (2013). Structural basis for Z-DNA binding and stabilization by the zebrafish Z-DNA dependent protein kinase PKZ. *Nucleic Acids Res.*, 41(21):9924–33. /cited on p. 96/

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, 66(Pt 4):486–501. /cited on p. 90/

Gjessing, M. C., Yutin, N., Tengs, T., Senkevich, T., Koonin, E., Rø nning, H. P., Alarcon, M., Ylving, S., Lie, K.-I., Saure, B., Tran, L., Moss, B., and Dale, O. B. (2015). Salmon Gill Poxvirus, the Deepest Representative of the Chordopoxvirinae. *J. Virol.*, 89(18):9348–67. /cited on p. 101/

Ilouze, M., Dishon, A., Kahan, T., and Kotler, M. (2006). Cyprinid herpes virus-3 (CyHV-3) bears genes of genetically distant large DNA viruses. *FEBS Lett.*, 580(18):4473–8. /cited on p. 88, 101/

Kabsch, W. (2010). XDS. *Acta Crystallogr. D. Biol. Crystallogr.*, 66(Pt 2):125–32. /cited on p. 89/

Kim, Y.-G., Muralinath, M., Brandt, T., Pearcy, M., Hauns, K., Lowenhaupt, K., Jacobs, B. L., and Rich, A. (2003). A role for Z-DNA binding in

vaccinia virus pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, 100(12):6974–9. /cited on p. 87/

Liu, T.-K., Zhang, Y.-B., Liu, Y., Sun, F., and Gui, J.-F. (2011). Cooperative roles of fish protein kinase containing Z-DNA binding domains and double-stranded RNA-dependent protein kinase in interferon-mediated antiviral response. *J. Virol.*, 85(23):12769–80. /cited on p. 87/

Michel, B., Fournier, G., Lieffrig, F., Costes, B., and Vanderplasschen, A. (2010). Cyprinid Herpesvirus 3. *Emerg. Infect. Dis.*, 16(12):1835–1843. /cited on p. 88/

Miwa, S., Ito, T., and Sano, M. (2007). Morphogenesis of koi herpesvirus observed by electron microscopy. *J. Fish Dis.*, 30(12):715–722. /cited on p. 88/

Perelberg, A., Smirnov, M., Hutoran, M., Diamant, A., Bejerano, Y., and Kotler, M. (2003). Epidemilogical Description Of A New Viral Disease Afflicting Cultured Cyprinus Carpio In Israel. *Isr. J. Aquac. - Bamidgeh*, 55(1):5–12. /cited on p. 87/

Placido, D., Brown, B. A., Lowenhaupt, K., Rich, A., and Athanasiadis, A. (2007). A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure*, 15(4):395–404. /cited on p. 86/

Rakus, K., Ouyang, P., Boutier, M., Ronsmans, M., Reschner, A., Vancsok, C., Jazowiecka-Rakus, J., and Vanderplasschen, A. (2013). Cyprinid herpesvirus 3: an interesting virus for applied and fundamental research. *Vet. Res.*, 44(1):85. /cited on p. 87/

Rothenburg, S., Deigendesch, N., Dittmar, K., Koch-Nolte, F., Haag, F., Lowenhaupt, K., and Rich, A. (2005). A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl. Acad. Sci. U. S. A.*, 102(5):1602–7. /cited on p. 87/

Schwartz, T. (1999). Crystal Structure of the Z Domain of the Human Editing Enzyme ADAR1 Bound to Left-Handed Z-DNA. *Science (80-.).*, 284(5421):1841–1845. /cited on p. 86/

Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, 282(5740):680–6. /cited on p. 86/

Yang, P.-J., Wu, C.-X., Li, W., Fan, L.-H., Lin, G., and Hu, C.-Y. (2011). Cloning and functional analysis of PKZ (PKR-like) from grass carp (Ctenopharyngodon idellus). *Fish Shellfish Immunol.*, 31(6):1173–8. /cited on p. 101/

3

Crystal structure of the Zα ORF112 with Z-DNA

Abstract

Protein kinase RNA-activated (PKR) is one of the key components of the innate immunity and is directed towards the recognition of pathogen double-stranded RNAs (dsRNA) in vertebrates. PKR is a modular protein which consists of two dsRNA binding domains (dsRBD) and a kinase domain. After detection of dsRNA by its dsRBDs, PKR dimerizes and autophosphorylates and this results in the activation of the kinase. The activated PKR then phosphorylates the eukaryotic initiation factor 2α (eIF2α) leading to inhibition of protein translation. Interestingly, some fish species (like zebrafish, koi and common carp) express a protein paralogous to PKR named PKZ, which instead of dsRBD domains contains two Z-DNA binding domains (Zα domains). As we have discussed in previous chapters these domains are found in other proteins such as ADAR1, DAI and poxviral E3L. Zαs are specialized domains with the ability to recognize the left-handed conformation of nucleic acids (both Z-RNA and Z-DNA). Recently, we have found that ORF112 from cyprinid herpesvirus 3 (CyHV3) contains a novel Zα domain. Intriguingly, ORF112 is encoded only in the genomes of viruses infecting fish species which have PKZ. Therefore, ORF112 can be a putative inhibitor of PKZ and may sequester the viral nucleic acids and prevent their recognition. We predict that Zα ORF112 can be an important player in the virus pathogenesis (similar Zα E3L). Previously, we solved the structure of the apoprotein confirming that Zα ORF112 is a bona fide Z-DNA/Z-RNA binding domain. Here, we report the structure of Zα ORF112 in complex with a $T(CG)_9$ oligonucleotide at 1.5 Å. We confirm that the DNA is in the left-handed conformation and in collaboration with the laboratory of Alain Vanderplasschen, we demonstrate that ORF112 can be targeted to stress granules (the stalled, pre-initiation translational complexes) like other Zα domains. These findings confirm that Zα ORF112 shares many similarities with other Z-DNA binding domains not only at the structural and biochemical but also at the functional level.

Publication

Krzysztof Kuś, Krzysztof Rakus, Maxime Boutier, Theokliti Tsigkri, Luisa Gabriel, Alain Vanderplasschen and Alekos Athanasiadis, The Structure of the Cyprinid Herpesvirus 3 ORF112-Zα·Z-DNA Complex Reveals a Mechanism of Nucleic Acids Recognition Conserved with E3L, a Poxvirus Inhibitor of Interferon Response, *The Journal of Biological Chemistry*, 2015, 290(52), p.30713-25

Contribution

Alekos Athanasiadis conceived the project. Alekos Athanasiadis and Krzysztof Kuś designed the experiments and analyzed the data. Krzysztof Kuś crystallized the ORF112 complex, collected data, determined the structure and performed ITC and band shift experiments. AV guided the experiments using CyHV3, Krzysztof Rakus and Maxime Boutier performed the CyHV3 experiments, Theokliti Tsigkri did ORF112 purification and Luisa Gabriel identified the Zα localization to SGs in human cells.

3.1 Introduction

In order to respond and counteract the viral infection, cells need to recognize features of the invading pathogen. Nucleic acids are among the molecules which can be detected leading to the activation of the cellular responses. One of the best-studied proteins involved in the recognition of dsRNA in the cytoplasm is PKR. This protein contains two dsRNA-binding domains (dsRBD) and a catalytic kinase domain. Upon binding to dsRNA, PKR undergoes structural changes that enable self-activation resulting in the phosphorylation of the eIF2α and translational suppression (Cole, 2007). In several bony fish species (such as zebrafish, common carp and koi), PKR coexists with a homolog named PKZ. PKZ,

instead of two dsRBD found in PKR, has two Zα domains (Rothenburg *et al.*, 2005). Zαs are currently the only domains known to interact and stabilize the left-handed DNA or RNA. These two kinases may respond to different nucleic acids triggering the repression of translation. It is thought that PKZ might have evolved to sense the transient left-handed conformation of nucleic acids. Interestingly, we have found that cyprinid herpesvirus 3 which belongs to the *Alloherpesviridae* family encodes ORF112 a protein with a Zα domain. As viruses of this family infect fish which express PKZ, ORF112 may be an inhibitor of the host interferon responses. Thus, Zα ORF112 may act to sequester viral nucleic preventing their detection by PKZ. Previously, such a role has been demonstrated for the poxviral Zα E3L and shown to be important for the pathogenesis of vaccinia virus in the mice model (Kim *et al.*, 2003). Zα E3L is thought to antagonize the DNA sensor DAI (a protein with two Zα domains).

In the previous chapter, we have provided evidence that free Zα ORF112 shares the same fold with other Z-DNA binding domains and binds $T(CG)_3$ in the left-handed conformation. The discovery of the Zα ORF112 parallels *Alloherpesviridae* (from *Herpesvirales* order) with poxviruses and may indicate a common ancestor or horizontal gene transfer between these viruses. As the details of the Zα PKZ (de Rosa *et al.*, 2013; Kim *et al.*, 2014) or E3L (Ha *et al.*, 2004) interaction with Z-DNA are known, we were motivated to understand the specificity and the mechanism of Z-DNA recognition by this novel Zα ORF112. In this chapter, we describe the structure of the ORF112·DNA complex and discuss its mode of binding to an 18 bp long Z-DNA. The crystal structure of the complex at 1.5 Å reveals that the functional unit interacting with the nucleic acid is an on-DNA Zα dimer. We propose that protein-protein interactions confer additional stability to the complex and block the Z-to-B transition (a conversion to the right-handed conformation). In this work, we used a $T(CG)_9$ oligonucleotide which is much longer than in previous studies, this allowed us to detect novel interactions that extend beyond the minimal $T(CG)_3$ binding site. We took advantage of the structural data to infer the relationship between Zαs from ORF112,

E3L and PKZ. Furthermore, we demonstrate that ORF112 is targeted to stress granules during oxidative stress, like the other members of Z-DNA binding family (Ng *et al.*, 2013).

3.2 Materials and Methods

Cloning, expression, purification

The Zα domain of ORF112 (BAF48926.1 residues 187-278) was cloned into a pET28a vector with NheI/XhoI restriction enzymes as a His-tag N-terminal fusion protein. The construct was expressed in the *Escherichia coli* strain BL21 (DE3). Cell cultures with 0.6-0.9 OD were induced with 0.7 mM IPTG. After 3h, cells were harvested by centrifugation (5000 g) at 4 °C. Chemical cell lysis was performed with BugBuster (Novagen) in the presence of 1 mM PMSF, a cocktail of proteinase inhibitors (Complete Mini, EDTA-free–Roche) and Benzonase (Novagen) for 1h at 4 °C. The protein extract was loaded on a HiTrap IMAC Sepharose FF column (GE Healthcare). The column was washed with 30 mM imidazole and then the protein was eluted using a gradient of 30-250 mM imidazole. The His-tag was cleaved with 10 U thrombin during an overnight dialysis at 4 °C against MonoS buffer A (10 mM HEPES pH 6.9, 20 mM NaCl) supplemented with 5 mM EDTA. The cleaved protein was loaded on a Mono S 4.6/100 PE (GE Healthcare). The column was washed with a gradient of 20-120 mM NaCl. The protein was eluted with 120-500 mM NaCl gradient and the content of the fractions was evaluated by gel electrophoresis. Buffer exchange and protein concentration was performed with Amicon-Ultra centrifugal filters (Merck-Millipore). The protein was concentrated to 35 mg/ml in 10 mM HEPES pH 7.4, 20 mM NaCl and used in crystallization trials.

Complex crystallization

A $T(CG)_9$ DNA oligonucleotide was purchased from Integrated DNA Technologies (IDT) and dissolved in MilliQ water. The oligonucleotides

were annealed overnight in a PCR machine using a temperature gradient from 80 °C to RT, decreasing 1 °C every 12 min. Protein and oligonucleotide concentration estimations were based on absorbance measurements with a NanoDrop device. For crystallization, a complex mix of Zα ORF112 (1.2 mM) with T$(CG)_9$ (0.3 mM) was incubated at 37 °C for 20 min and screened against solutions of 3D structure Screen (Molecular Dimensions). Initial small hexagonal crystals were obtained in 1.05 M lithium sulfate, 0.1 M HEPES pH 7.5. The crystals were optimized and the best quality ones were obtained in 0.9 M lithium sulfate, 0.1 M HEPES pH 7.6. Such crystals were harvested and cryo-protected either in 20% glycerol, 20% PEG200 or Paratone-N and flash-frozen in liquid nitrogen.

Data collection, structure determination and phylogenetic analysis

X-Ray diffraction data of crystals frozen in liquid nitrogen were collected at the ID29 beamline of ESRF/Grenoble synchrotron at 100 K using 0.976 Å x-ray wavelength. The best quality data were obtained from crystals cryo-protected in 20% PEG200 (immersed in mother liquor with 20 % PEG200 for 2 min). The XDS package was used to process the data (Kabsch, 2010). The complex crystallized in the $P3_221$ space group with unit cell dimensions and angles a=44.82 Å, b=44.82 Å, c=140.08 Å, α=β=90.00°, γ=120.00°. Initial phases were obtained by molecular replacement using a composite model of a truncated chain (48 residues of 69) of Zα ORF112 (PDB ID: 4HOB) with 5 bases of a Z-DNA strand from the Zα ADAR1 DNA complex (PDB ID: 1QBJ). The composite model was constructed after superposition of the two structures in PyMol. Our starting model was then refined in Phenix (Adams *et al.*, 2010) followed by manual rebuilding in Coot (Emsley *et al.*, 2010) to a final R/Rfree of 0.18/0.21. The final model has phi-psi angles for all protein residues within the favored region of the Ramachandran plot. The asymmetric unit of the crystals contains two Zα domains and two Z-DNA chains of 6 bases (one-third of the T$(CG)_9$). The overhanging T is disordered

and not visible in the electron density. Electron density for the DNA shows continuity between asymmetric units and neighboring unit cells forming infinite helices spanning the lattice in three directions. The terminal phosphates of the DNA backbone in the pseudo-continuous helices are not aligned among them and this explains the disappearance of the DNA ends and leads to an apparent asymmetric unit that contains only a third of the physical DNA molecule. Similar cases have been observed in the crystal structures of Z-DNA when the helical axis coincided with a crystallographic axis (Brennan *et al.*, 1986; Ban *et al.*, 1996; de Rosa *et al.*, 2013). In the final model, the density of two solvent exposed residues (Lys233A, Gln270A) was very weak and thus these side-chains were not modeled. In addition, the N-terminal 36 residues (six originating from the expression vector) as well as the last C-terminal residue are not seen in the electron density and thus are not modeled. Arg258B also did not show a strong density but guided by the interaction of the same residue in chain A, we were able to model this residue in two alternative conformational states. The final model contains 120 waters as well as four sulfate ions bound to highly charged ORF112 surfaces. Details about data collection and refinement statistics are listed in **Table A.2** included in **Appendix**. The PISA (Krissinel and Henrick, 2007) software was used to obtain information about DNA-protein, protein-protein interfaces and assemblies. Structure-guided alignments were performed in UCSF Chimera (Pettersen *et al.*, 2004) and phylogenetic analysis in PhyML (Guindon and Gascuel, 2003) as implemented on the publicly available server Phylogeny.fr (Dereeper *et al.*, 2008). Representations of the structure and structural alignments were generated in PYMOL (Delano, 2002). The RMSDs from the structural alignments refer to chain A from each structure with the exception of drPKZ were chain B was used, as chain A has a disordered region. The model and structure factors have been deposited to the RCSB database (PDB ID: 4WCG)

Gel-Shift assay

In order to compare the binding affinity of Zα ORF112 and Zα

ADAR1, we employed the gel mobility shift assay. Mixtures of proteins with 20 µM $T(CG)_8$ or $T(TG)_8$ oligonucleotides were prepared at 2/1 and 4/1 ratios and incubated at 37 °C for 20 min. Next, the samples were separated on a non-denaturing polyacrylamide gel (6% DNA Retardation Gels, ThermoFisher) and stained for nucleic acids (RedSafe) followed by protein staining (Coomassie blue).

Cell culture, virus and treatment

Common Carp Brain (CCB) cells (Neukirch *et al.*, 1999) were cultured in minimum essential medium (Sigma) containing 4.5 g/L glucose (D-glucose monohydrate; Merck) and 10% fetal calf serum (FCS) as described previously (Costes *et al.*, 2008). The CyHV3 FL strain was isolated from the kidney of a fish that died from CyHV3 infection (Costes *et al.*, 2008). To induce stress granule formation CCB cells were incubated at 25 °C for 30 min in media supplemented with 1 mM sodium arsenite (Sigma), then washed twice with complete medium and allowed to recover for 15 min before further processing. Purification of CyHV3 virions (American strain: accession code: ABG42939.1) and mass spectrometry analyses by 2D LC-MS/MS were performed as described previously (Michel *et al.*, 2010).

Immunofluorescent staining and confocal microscopy

CCB cells were fixed in PBS containing 4% (w/v) paraformaldehyde for 15 min at 4 °C and then 10 min at 20 °C. After washing with PBS, samples were permeabilized in PBS containing 0.2% (w/v) Triton-X100 at 20 °C for 10 min. Immunofluorescent staining (incubation and washes) was performed in PBS containing 10% FCS (v/v). CCB cells were incubated at 37 °C for 60 min with mouse polyclonal sera raised against CyHV3 ORF112 protein and rabbit polyclonal antibodies raised against CyHV3 purified virions or rabbit polyclonal antibodies raised against HuR/ELAVL1 protein (Proteintech). After three washes, samples were incubated at 37 °C for 30 min with Alexa Fluor 488-conjugated goat anti-mouse IgG (H+L) (LifeTechnologies) and with Alexa Fluor 568-conjugated

goat anti-rabbit IgG (H+L) (LifeTechnologies) as the secondary antibodies. After washing, cells were mounted using Prolong Gold Antifade Reagent with DAPI (Invitrogen).

Isothermal titration calorimetry

Binding heat from the complex formation between Zα domains and DNA was measured on an ITC200 instrument (GE Healthcare) at 25 °C, 1000 rpm. Oligonucleotides T$(CG)_3$ and T$(CG)_6$ were purchased from IDT and annealed. Protein and DNA storage solutions were exchanged against 10 mM HEPES pH 6.5, 50 mM NaCl with Amicon Ultra Centrifugal Filters (Merck Millipore). Briefly, experiments consisted of 18 injections of 2 µl of protein to oligonucleotide (concentrations used were optimized for optimal curve fit and are indicated in the corresponding figure legends). After each injection, the system was allowed to equilibrate for 3 min. Raw data were integrated using NITPIC software (Keller *et al.*, 2012), fitting was carried out with SEDPHAT (one site models: Levenberg–Marquardt algorithm) (Houtman *et al.*, 2007) or in CHASM (two site models) (Le *et al.*, 2013). Plots were created with GUSSI (evoked in SEDPHAT). As negative controls, we performed injection of the proteins to the buffer and we confirmed that heat of injection was constant (data not shown). Thermodynamic parameters for binding of Zα domains to T$(CG)_n$ oligonucleotides are summarized in **Table A.3** in **Appendix**.

3.3 Results

Multiple trials aiming to obtain crystals Zα ORF112 and T$(CG)_3$ complex were unsuccessful. We decided to extend the Zα domain construct to start from M187 (29 amino acids more than the protein used in experiments described in **Chapter 2**). Additionally, the crystal structure of Zα PKZ with T$(CG)_6$ solved in our laboratory indicated that the interaction with nucleic acids can extend beyond minimal $(CG)_3$ duplex (de Rosa *et al.*, 2013). Based on this knowledge, we tested

different oligonucleotide lengths in the cocrystallization experiments and successfully we obtained crystals with $T(CG)_7$ (**Figure 3.1A**) and $T(CG)_9$ (**Figure 3.1B**) but only the latter diffracted to high resolution.

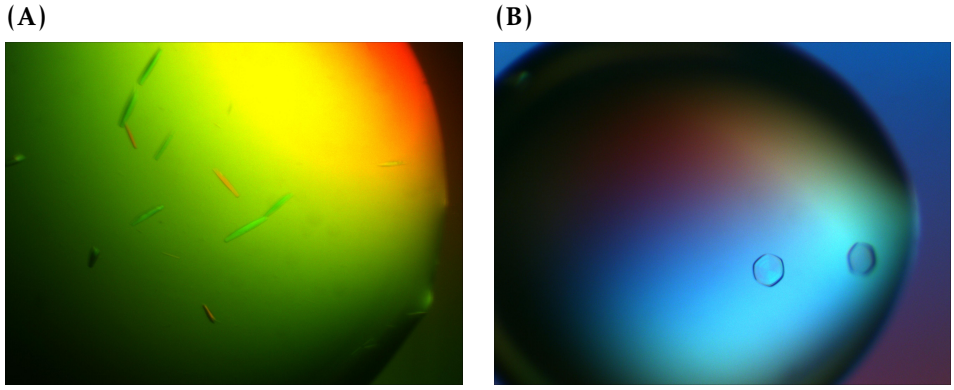


Figure 3.1: Crystals of Zα ORF112 with Z-DNA. (A) Crystals with $T(CG)_7$. Condition: 26 % PEG4000, 0.2 M sodium acetate, 0.1 M Tris pH 8.5, 170 mM guanidine hydrochloride, 4.5 mM nicotinamide adenine dinucleotide hydrate. No significant diffraction. **(B)** Crystals with $T(CG)_9$. Condition: 0.90 M lithium sulfate, 0.1 M HEPES pH 7.6.

3.3.1 General description of Zα ORF112·Z-DNA complex structure

The X-ray diffraction pattern of Zα ORF112·$T(CG)_9$ revealed that the crystal belongs to the $P3_221$ space group. We were able to solve the structure of the complex at 1.5 Å obtaining initial phases using the molecular replacement (see **Materials and Methods**). The asymmetric unit is composed of two protein monomers and two six bp long DNA strands (**Figure 3.2A**). As the DNA forms pseudo-continuous helices throughout the crystal lattice, the DNA bases in the asymmetric unit were chosen to capture the maximal number of interactions between protein and DNA without a need to apply symmetry operations (**Figure 3.2A**). The structure confirmed that Zα ORF112 belongs to the wHTH family in agreement with the findings on the free protein. The structural comparison between the free and DNA-bound form shows that the domains adopt very similar conformation with the exception of the

swapped segment (0.25-0.36 Å RMSD depending on the choice of the chains, excluding the region of domain-swapping). Generally, Zα domains do not share high sequence identity but are conserved on the structural level. It is not surprising that Zα ORF112 shares the fold with other domains of this family. However, differences found in the wing region are of great interest (**Figure 3.2B**).

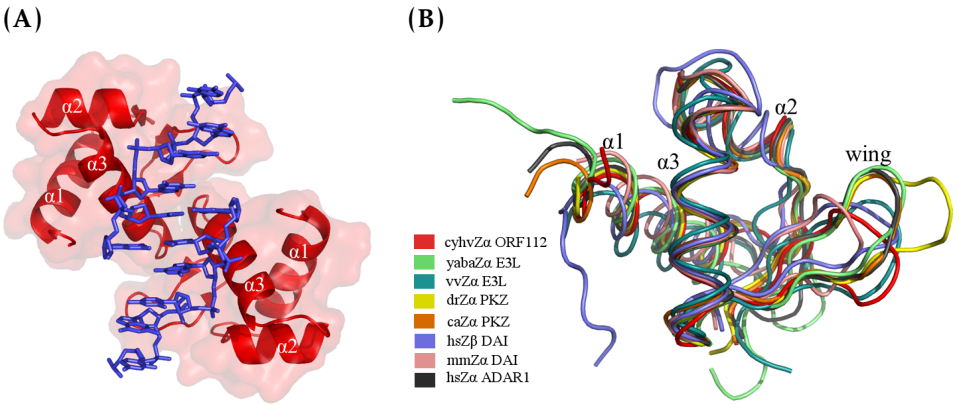


Figure 3.2: The content of the asymmetric unit of Zα ORF112 complexed with Z-DNA and its structural similarity to other members of Zα family. (A) The asymmetric unit content representation of the cyprinid herpesvirus 3 Zα ORF112 with the left-handed DNA: two protein domains (depicted as red cartoon with a semi-transparent surface) and two Z-DNA (6 bp) strands (as blue stick models) **(B)** The structural comparison (based on Cα atoms alignment) of Zα family members with known structures: cyprinid herpesvirus 3 Zα ORF112 (cyhv3Zα ORF112, PDB ID: 4WCG), Yaba-like disease virus Zα E3L (yabaZα E3L, PDB ID: 1SFU), vaccinia virus Zα E3L (vvZα E3L, PDB ID: 1OYI), *Danio rerio* Zα PKZ (drZα PKZ, PDB ID: 4LB5), *Carassius auratus* Zα PKZ (caZα PKZ, PDB ID: 4KMF), *Homo sapiens* Zα2 DAI (hsZα2 DAI, PDB ID: 3EYI), *Mus musculus* Zα DAI (mmZα DAI, PDB ID: 1J75), *Homo sapiens* Zα ADAR1 (hsZα ADAR1, PDB ID: 1QBJ)

As previously mentioned, the pseudo-continuous DNA helices span throughout the crystal and this enables to determine how multiple Zα molecules are arranged along the Z-DNA helix. For this purpose, we recreated a helix corresponding to $(CG)_9$ (**Figure 3.3A** (top view) and **B** (along DNA axis view)) and defined the arrangement of Zα ORF112 by taking as a reference the CH-π interaction between Tyr257 and guanosine in syn conformation. As depicted in (**Figure 3.3C**), Zα ORF112 binding sites are separated by six bp on the same strands while, binding to the

complementary strand is shifted by three bases. This type of arrangement was previously found only for Zα PKZ with T$(CG)_6$ which contained also pseudo-continuous helices. In the case of Zα ORF112, this type of spacing is possible because monomers are engaged in the symmetric protein-protein contacts which involve $\alpha 3$ (helix 3) of one chain and the wing region of the other molecule. In the following paragraph will discuss these interactions in detail.

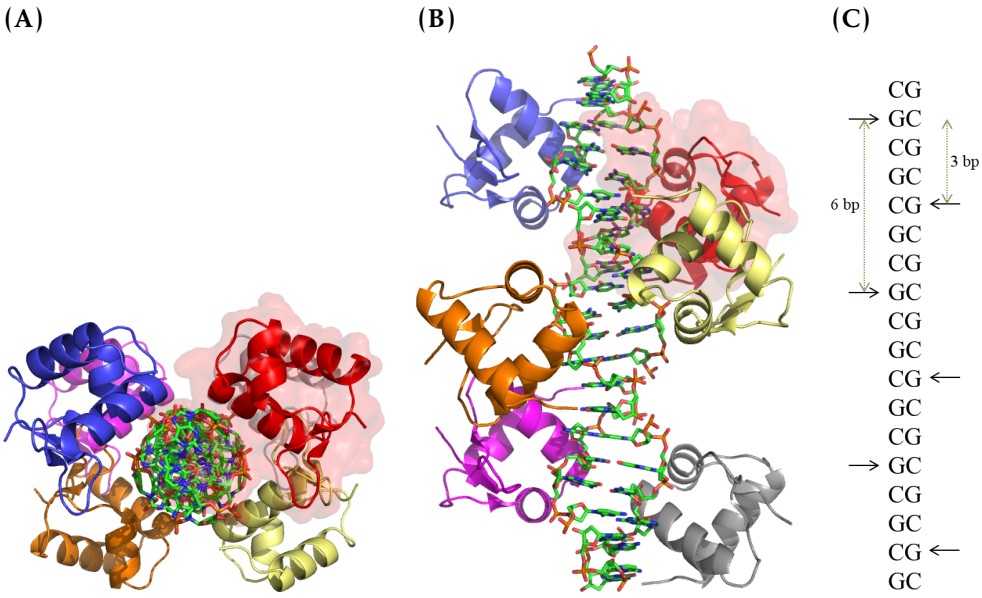


Figure 3.3: The arrangement of Zα ORF112 on long Z-DNA. (A) and **(B)** show the recreated biological assembly of $(CG)_9$ with bound Zα ORF112 domains. The view is perpendicular to (A) or along (B) the DNA axis. Z-DNA is presented as a stick model. Each of the protein monomers is shown as a ribbon with a different color. The red semi-transparent surface, marking one of the monomers, serves as a reference. Six monomers decorate 18 bp Z-DNA. **(C)** Diagram of the arrangement of Zα ORF112 domains along reconstructed Z-DNA. The characteristic CH–π bond between Tyr257 and the guanosine in the syn conformation is used as the reference and is depicted by the horizontal black arrows. The vertical arrows (dotted, brown) indicate a number of base pairs between adjacent monomers (the binding site spacing).

3.3.2 On-DNA Zα ORF112 dimerization

In almost all previous crystal structures of the Zα with nucleic acids a 6 bp $T(CG)_3$ segment was used as a ligand. These structures revealed that two monomers bind to the complementary strands of the DNA helix and no interactions between proteins could be seen. Based on such structures it was postulated that Zαs bind each DNA strand independently. This view has changed with a previous study performed in our laboratory which involved longer Z-DNA fragments. The structure of *Danio rerio* Zα with $T(CG)_6$ indicated that protein-protein contacts are possible and proteins can interact with the opposite DNA strand (de Rosa *et al.*, 2013). In addition, the structure of the free Zα ORF112 uncovered a novel protein-protein interaction (domain-swapping) which suggested that dimerization may be important in the nucleic acid binding.

The structure of the Zα ORF112 complex with Z-DNA unveiled protein dimer formation upon DNA binding and this on-DNA dimer is mostly stabilized by van der Waals interactions. This dimer is symmetrical as both monomers are related by a two-fold non-crystallographic symmetry (**Figure 3.4A, B**). The C-terminal part of the α3 helix of one monomer interacts with the β2 strand of the second Zα. This reciprocal interaction provides stabilization and anchors the position of the wing region. Seven residues of each monomer (10 % of the total surface area) contribute to the dimer stabilization. Two hydrogen bonds are formed between Lys267 (N-H) of one chain and Glu261 (C=O) of the other monomer. Additionally, a sulfate ion forms salt bridges between Arg266 of each of the monomers (**Figure 3.4B**). This protein-protein interface is extensive but formed only in the presence of DNA as we have previously shown by size exclusion chromatography that in the absence of sulfate ions the protein is monomeric (refer to **Chapter 2**).

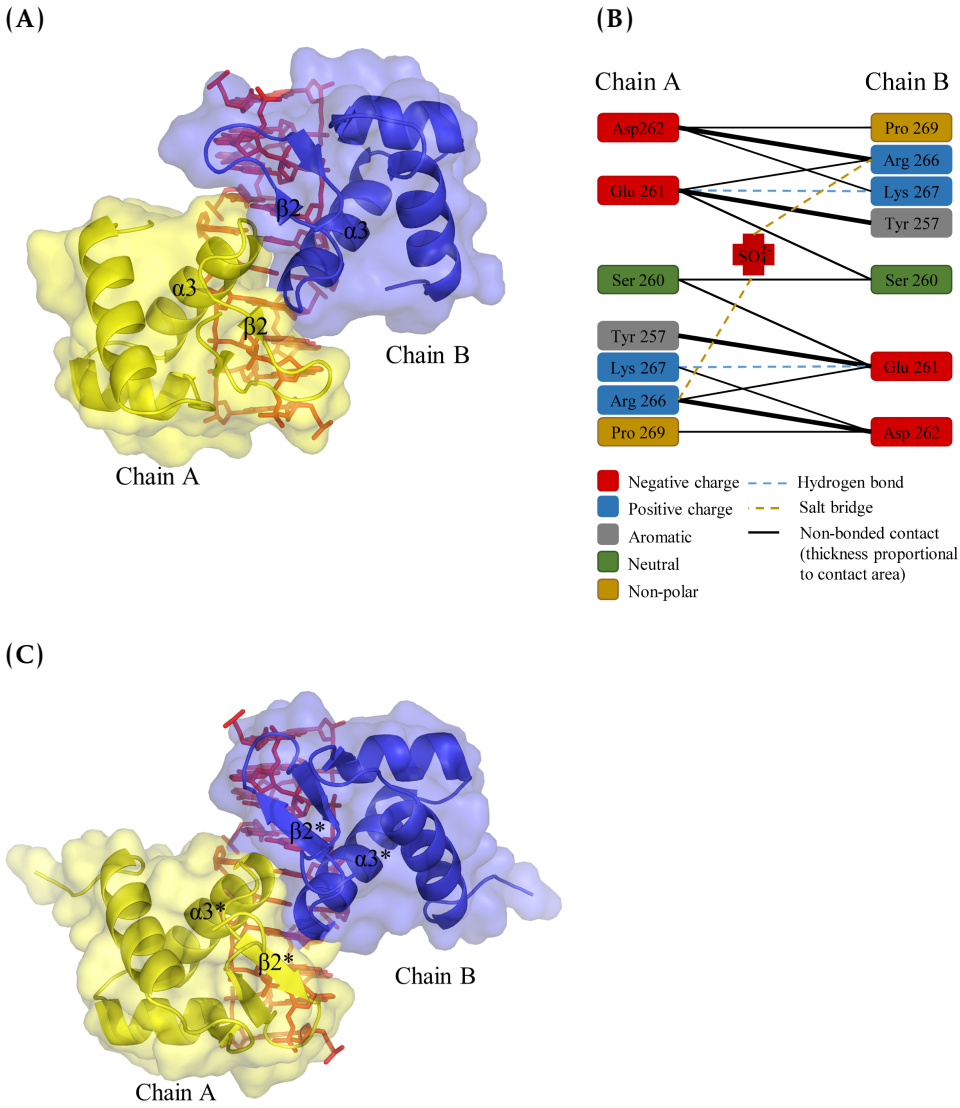


Figure 3.4: The DNA-mediated protein-protein interactions between Zα ORF112 domains. (A) A schematic representation of protein-protein interface between CyHV3 Zα ORF112 domains (protein shown as a blue and yellow cartoon with a semi-transparent surface). Z-DNA is 10 bp long (a red stick model) and was reconstructed from the asymmetric unit by symmetry operation. **(B)** A summary of symmetrical interactions between Zα ORF112 domains. Presented hydrogen bonds are formed between the backbones of the monomers. **(C)** A putative complex of Yaba-like disease virus Zα E3L (model based on the Zα ORF112 complex). Monomers of Zα E3L can form on-DNA dimer without clashes between the backbone atoms of the monomers.

As mentioned in the previous chapter, the sulfate ion in the apoprotein occupies the exact same position of the phosphate group of Z-DNA based on superimposition with Zα ADAR1. These findings were confirmed in the structure of Zα ORF112·DNA complex and moreover, the critical three residues involved in the interaction with Z-DNA are found in exactly the same conformations in apoprotein·sulfate and protein·DNA complex (**Figure 3.5**).

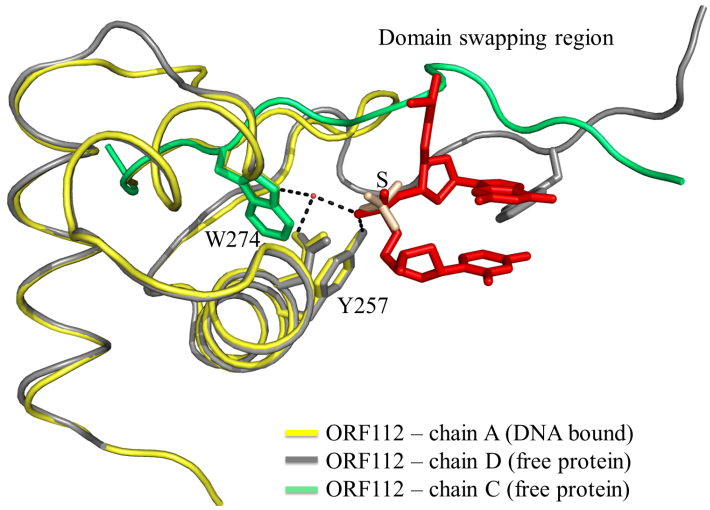


Figure 3.5: Superimposition of the free and Z-DNA bound Zα ORF112 structures. Protein chains are shown as cartoon loops: Zα bound to the left-handed DNA (yellow), two chains from free protein dimer (gray and green). For the sake of clarity, most of the chain C (green) is omitted except for the region involved in the domain-swapping. For the same reason, only two bases of the DNA are shown as a red stick model. The sulfate ion (marked as S, wheat stick model) from the apoprotein structure is shown to occupy the same position as the phosphate from the Z-DNA backbone.

The crystal structure of PKZ Zα indicated that locking the DNA molecule in the Z-conformation might require more than one Z-DNA binding domain (de Rosa *et al.*, 2013). In agreement, most of the proteins with Zα domains contain two (or three in case *S. purpuratus* ADAR1) repetitions of this domain (**Figure 1.11**). Interestingly, the poxviral E3L and ORF112 (cyprinid herpesvirus 2, 3) have only one copy of the Zα domain and the formation of on-DNA dimer we found may provide the equivalent stabilization of the left-handed conformation. To gain insight into whether poxviral (Yaba-like disease virus) Zα E3L may also form a

similar dimer on Z-DNA, we superimposed Zα E3L domains on the Zα ORF112. As depicted in **Figure 3.4C**, Zα E3L could indeed assemble the on-DNA dimer without any clashes of the main backbone, with only a few side-chain overlaps (but they could adopt alternative conformations). Moreover, the C-terminal part of the Zα E3L (the dsRBD) would not interfere with such dimer assembly. Overall, we conclude that the viral Zαs might utilize the on-DNA dimerization as a mean to maintain the Z-conformation (blocking the transition back to B- or A- conformation).

3.3.3 Unique features of Zα ORF112 bound to Z-DNA

As mentioned before, Zα domains recognize the characteristic zig-zag shape of the left-handed nucleic acid conformation. The interactions are based on a network of protein contacts with the sugar-phosphate backbone of the Z-conformation and a unique CH-π bond between a critically conserved tyrosine and a guanosine (purine) in the syn conformation. Additionally, the asparagine and tryptophan are part of the core of the recognition mechanism. This triad of residues is located in the α3 helix and the wing region of the protein. In this respect, Zα ORF112 is similar to other Z-DNA binding domains as these three critical amino acids are conserved (Tyr257, Asn253 and Trp274). In this section, we compare Z-DNA binding by Zα ORF112 and Zα E3L. Additionally, we will describe the novel features of Zα ORF112 bound to the left-handed DNA.

Differences between viral Zα domains

Given a potential link between poxviruses and cyprinid herpesviruses, we decided to compare binding modes of Zα domains from different virus families (**Figure 3.6**). The most profound differences in the Z-DNA binding by the cyprinid herpesvirus Zα ORF112 and Yaba-like disease virus Zα E3L are found in the wing region. Zα ORF112, unlike other Zα domains, has the wing region pointing away from Z-DNA. The only

contacts between the nucleic acid and protein are formed by the conserved Trp274 and Pro272 (in the γ-turn) (**Figure 3.6A, C**).

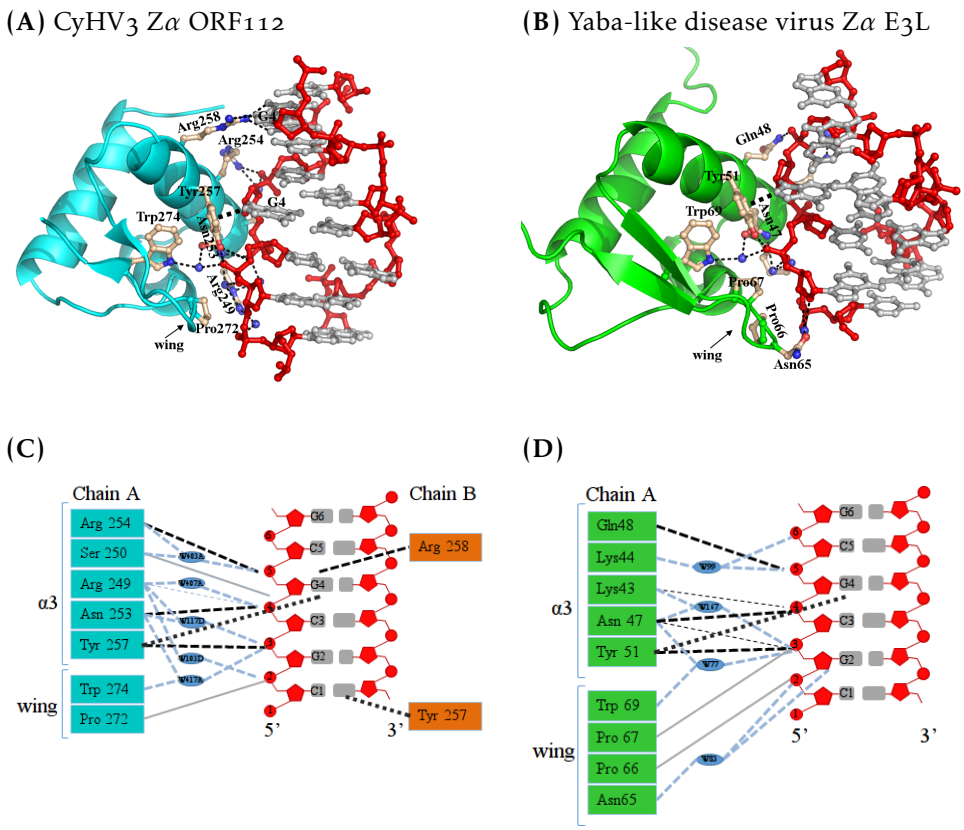


Figure 3.6: Comparison of DNA binding by the viral Zα domains from ORF112 (cyprinid herpesvirus 3) and E3L (Yaba-like disease virus). (A) The interaction between CyHV3 Zα ORF112 and a reconstructed 7 bp Z-DNA duplex. The protein chain is depicted as a cyan ribbon and the residues involved in the interaction with Z-DNA are shown as a ball-and-stick model (wheat color - carbon, blue - nitrogen, red - oxygen). The phosphate backbone and sugars are colored red, and bases gray. Important waters are represented as blue spheres. **(B)** The Z-DNA binding interface of Yaba-like disease virus Zα E3L (6 bp, PDB ID: 1SFU). The protein is shown as a green cartoon; all other depictions are as described in (A). **(C)** and **(D)** Diagram summarizing the interactions between Zα ORF112 (C) and Yaba-like disease virus Zα E3L (D) with a 6 bp Z-DNA. The phosphate backbone and sugars are colored red, bases are gray. The direct hydrogen bonds are represented as black dashed lines and water-mediated bonds are shown as light blue dashed lines. Non-bonded contacts are shown as solid light gray lines. The characteristic CH-π bond between tyrosine and guanosine in the syn conformation is drawn as a dotted black line. Blue ovals denote the water molecules.

In contrast, the Zα E3L wing region has four amino acids interacting with the Z-DNA backbone, either by van der Waals or water-mediated hydrogen bonding (**Figure 3.6B, D**). Overall, the wing region of Zα domains has a significant degree of flexibility. As previous structural studies used a T$(CG)_3$ oligonucleotide, the wing region was in contact with the terminal base pair which is slightly distorted relatively to the internal bases. Therefore, the positioning of the wing region may be imposed by the Z-DNA backbone. Additionally, in the Zα ORF112 structure the two monomers form an on-DNA dimer stabilizing the wing region positioning. The C-terminal part of the recognition $\alpha 3$ helix of one chain interacts with $\beta 2$-strand of another monomer. Hence, other Zα domains might change the wing orientation and engage in the similar protein-protein interactions observed for Zα ORF112 if bound to longer Z-DNA fragments.

The $\alpha 3$ recognition helix of Zα domains is less flexible than the wing region. Two critical residues involved in the Z-DNA recognition are found in the same conformation: Tyr257, Asn253 for Zα ORF112 and Tyr51, Asn47 for Zα E3L. Other amino acids contributing to the Z-DNA binding differ between viral Zα domains. Zα ORF112 uses three positively charged arginine residues (Arg249, Arg254 and Arg258) to bind the Z-DNA backbone. Unexpectedly, Arg258 also establishes two direct hydrogen bonds with the guanine base (**Figure 3.7**). These direct base contacts (which suggest a sequence specific recognition) take place on the complementary strand (the opposite strand to the one with all other protein-DNA interactions) and have not been observed for any other Zα domain before. Interestingly, the base contacted by Arg258 is the same guanosine which forms a CH-π bond with Tyr257 of the second protein chain (**Figure 3.7**). This arrangement does not alter the conformation of Tyr257. This novel interaction may provide further stabilization of Z-conformation, impacting the Z-to-B transition. Although only some Zα domains have arginine in the equivalent position (like Zα domains of DAI), Yaba-like disease virus Zα E3L could use lysine for a similar purpose. This base-specific interaction occurs outside the minimal $(CG)_3$ substrate

and was only captured because our crystal structure is based on a longer Z-DNA fragment. Arg249 is involved in the network of water-mediated hydrogen bonds and Arg254 (the functional equivalent of Gln48 in Zα E3L) forms a direct hydrogen bond with a phosphate group from DNA backbone. Finally, Ser250 Zα ORF112 (found in an equivalent position in Zα PKZ) contributes a water-mediated hydrogen bond and van der Waals interactions to the Z-DNA phosphate backbone.

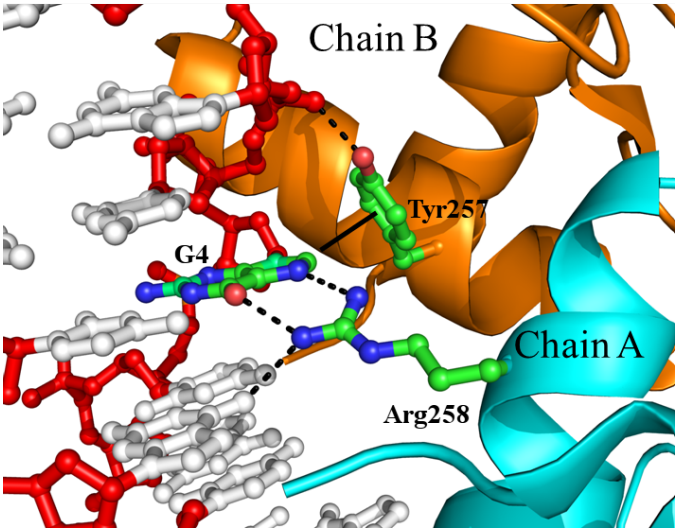


Figure 3.7: Direct base contacts with Z-DNA formed by two Zα ORF112 monomers. The phosphate and sugar backbone of DNA are shown as a red ball-and-stick model. The side chains of Arg258 (chain A) and Tyr257 (chain B) make contacts with the guanosine in the syn conformation (G4) - elemental coloring. Dotted, black lines depict the hydrogen bonds and a solid, black line indicates the characteristic CH-π bond between tyrosine and guanosine.

In order to understand the structural similarities between Zα ORF112 and Zα domains from PKZ and poxviruses, we performed structural alignments of Zα ORF112 with *Danio rerio* Zα PKZ, *Carassius auratus* Zα PKZ, Yaba-like disease virus Zα E3L (yabaZα E3L), vaccinia virus Zα E3L (refer **Figure 3.2B**). We observed that, structurally, *Danio rerio* Zα PKZ is the most similar to the Zα ORF112 (yielding 0.502 Å RMSD based on Cα atoms). Subsequently, we created a structure-guided sequence alignment (**Figure 3.8A**) to gain insight into residues which are equivalent between Zα domains. In addition, we aligned several sequences (Muscle software) and this served as an input for phylogenetic analysis (**Figure 3.8B**).

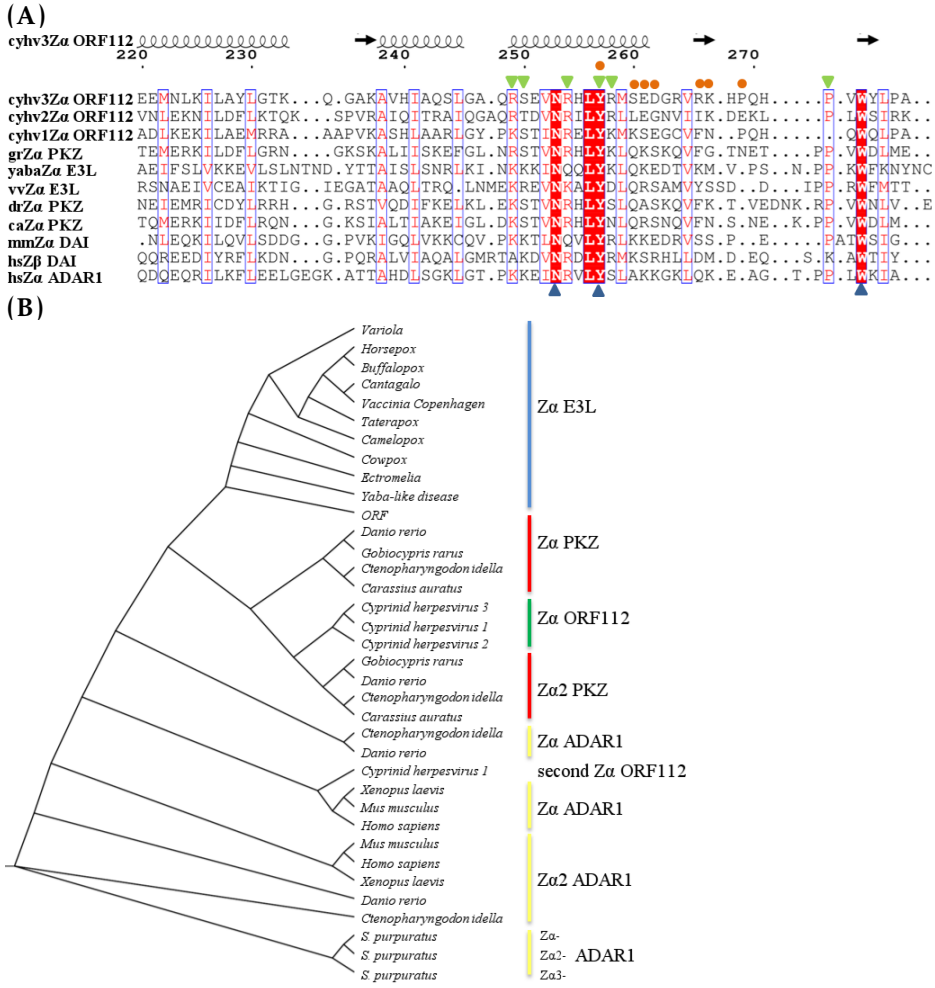


Figure 3.8: The phylogenetic analysis of Zα ORF112. (A) The structure-guided alignment of cyprinid herpesvirus 3 Zα ORF112 with other members of the Zα family (for PDB codes refer to **Figure 3.2A**). Additional three sequences without structural data are included: cyprinid herpesvirus 1 Zα ORF112 (cyhv1Zα ORF112), cyprinid herpesvirus 2 Zα ORF112 (cyhv2Zα ORF112), *Gobiocypris rarus* Zα PKZ (grZα PKZ). The top of the alignment contains the schematic representation of Zα ORF112 secondary structure (cyhv3Zα ORF112). The amino acids involved in the protein/DNA (green triangles) and protein/protein interactions (orange circles) are indicated above the alignment. The blue triangles below the alignment denote the three critical Z-DNA/Z-RNA binding residues – Asn, Tyr and Trp. The blue boxes surround positions with conservation higher than 50% and red shadings highlight absolute conservation in this alignment. **(B)** Cladogram of Zα domain family generated based on curated (no gaps) alignment (Muscle software) with PhyML using Jones-Taylor-Thornton (JTT) substitution model with SH-aLRT (Shimodaira–Hasegawa approximate likelihood ratio test).

Although based on the cladogram we cannot resolve the full history of Zα domain evolution, we observe that Zα domains from all known cyprinid herpesviruses cluster together and they are related to the Zα domains from PKZ (**Figure 3.8B**). Moreover, our analysis suggests that Zαs E3L form another group, which separated earlier than Zα ORF112 and PKZ. The similar conclusions are drawn from the structure-guided alignment where we included *Gobiocypris rarus* Zα PKZ being the most similar to cyprinid herpesvirus 3 Zα ORF112 (**Figure 3.8A**). Therefore, Zα ORF112 could have originated from Zα PKZ (co-opted from the host by the ancestor of cyprinid herpesviruses). Although we cannot exclude a horizontal gene transfer or a common ancestor with poxviruses and that the structural similarities between Zα ORF112 and PKZ have their basis in the high selective pressure as they may target similar nucleic acids (convergent evolution).

Distinct modes of Z-DNA binding by Zα ORF112 and Zα ADAR1

In order to evaluate the predictions concerning on-DNA dimerization obtained from the structure of Zα ORF112 with Z-DNA, we generated corresponding mutants of the domain and compared their binding affinities with that of the wild-type (WT) domain. In addition, we wanted to compare affinities of WT Zα ORF112 and Zα ADAR1 in the context of different lengths of Z-DNA. Among the available methodologies, we opted for isothermal titration calorimetry (ITC) as it allows measuring a binding affinity without a need to modify/immobilize binding partners. One of the requirements of ITC is a precise measure of the concentration of components but in our experiments, we need to consider that one of the components is the result of a B-to-Z equilibrium which complicates the analysis. When a cell of the instrument is filled with the oligonucleotide and only a fraction of the nucleic acid is in the Z-conformation. Each injection leads to Z-DNA binding and shifts the equilibrium towards the left-handed form. Hence, although we can provide the concentrations of DNA, we do not have any means to supply reliable estimates of

Z-DNA concentrations. If we assume that entire amount of DNA is in Z-form, the experiments provide underestimates of stoichiometries as we overestimate the Z-DNA concentration. Unfortunately, as there is no model that would incorporate the B-to-Z transition, we interpreted the data with one or two site models. Thus, in interpreting our results one should be aware that presented dissociation constants (inverse of the binding constant) are indicative of relative binding affinities of different proteins but most likely poor estimates of the absolute affinities (**Figure 3.9** and **3.10**).

Our ITC experiments revealed that WT Zα ORF112 binds the longer oligonucleotide $T(CG)_6$ (K_d=263 nM) with the affinity two times higher than the short $T(CG)_3$ (K_d=538 nM) suggesting that the extension of the minimal binding site contributes to stronger binding. Interestingly, thermodynamics of Zα ORF112 binding to $T(CG)_3$ and $T(CG)_6$ differ. Binding to the longer oligonucleotide is associated with the unfavorable entropy change ($-T\Delta S$=22.2 kJ/mol) which contrasts with the favorable entropy component of the binding to $T(CG)_3$ ($-T\Delta S$=-6.1 kJ/mol). The entropic penalty most likely reflects an ordering of the flexible protein regions (**Table A.3** in **Appendix**). Under the same conditions, Zα ADAR1 binds $T(CG)_3$ slightly better than WT Zα ORF112 but it should be noted that the one-site model does not fit well the points of the early injections. Interestingly, Zα ADAR1 binding to $T(CG)_6$ exhibits a different binding mode than WT Zα ORF112. The best agreement between model and the Zα ADAR1 data was obtained by applying the two binding sites model with K_{d1}=210 nM, and K_{d2}=8.7 µM (**Figure 3.9**). One might interpret these results in the view of Zα ORF112 on-DNA dimerization which may not happen for ADAR1. On the other hand, Zα ADAR1 may interact with a minimal fragment $(CG)_3$ of the repeat (with B-Z junction present) and the following second binding would be formed with the rest of the oligonucleotide.

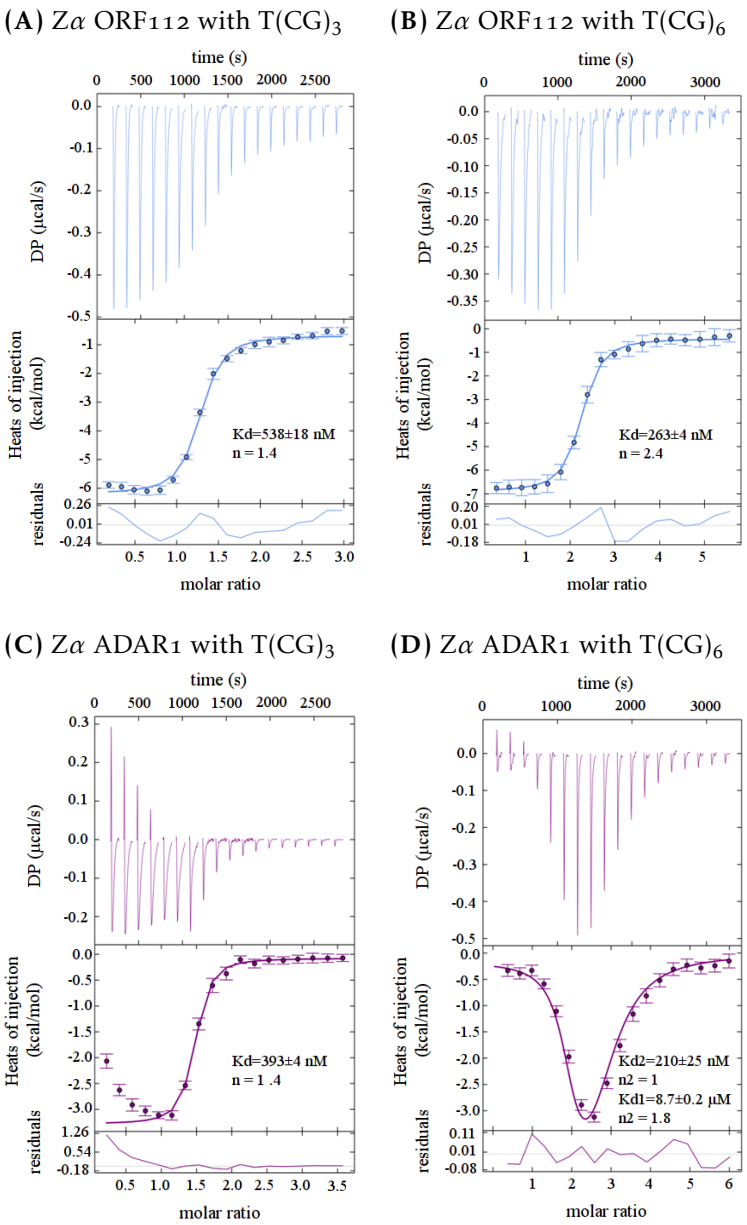


Figure 3.9: Isothermal titration calorimetry analysis of Z-DNA binding by Zα ORF112 and ADAR1. The isothermal titration calorimetry experiments were performed to determine the binding affinity of Zα ORF112 against $T(CG)_3$ (675 µM protein in the syringe against 45 µM oligonucleotide in the cell) **(A)** and $T(CG)_6$ (600 µM protein against 20 µM oligo) **(B)**. The same experiments for Zα ADAR1 are shown in panels: **(C)** (1200 µM protein against 70 µM oligo) and **(D)** (1200 µM protein against 40 µM oligo). All experimental data were fitted with a one-site model, except for ADAR1 against $T(CG)_6$ where a two-binding site model was used.

As we discussed in the previous section, two Zα ORF112 monomers engage in DNA-mediated protein-protein interactions. This on-DNA dimer may provide additional stabilization of the complex. In order to evaluate this possibility, we created a mutant Zα ORF112 by changing Ser260 either to Leu or Glu. Ser260 is a residue in the terminal region of the α3 helix which is at the interface of the protein-protein interaction. A substitution of serine for leucine (S260L) should interfere with dimer formation due to steric hindrance, whereas the other mutation (S260E) should additionally lead to the repulsion between monomers. In agreement with our predictions, we observe that S260L Zα ORF112 binds $T(CG)_6$ with lower affinity than WT Zα ORF112 (K_d= 470 nM versus 263 nM, **Figure 3.10A**). The entropic penalty of $T(CG)_6$ binding by S260L Zα ORF112 is slightly higher than for WT Zα ORF112 (-TΔS = 22.2 kJ/mol versus 25.8 kJ/mol, **Table A.3** in **Appendix**). This increase in the entropic cost is not fully compensated by the decrease of the enthalpy component. It is possible that a bulky leucine residue induced rearrangements that have a higher entropic penalty. For S260E Zα ORF112 the change in the affinity was far more dramatic (we could not assign a reliable value, data not shown). However, in this case, we cannot rule out that what we observe is a result of the repulsion of the DNA backbone by the glutamic acid and not solely the inability to form the on-DNA dimer. Moreover, as the structure of Zα ORF112 revealed the second strand interactions mediated by Arg258, we tested R258A Zα ORF112 against $T(CG)_6$ and we observed a moderate reduction in the affinity compared to WT Zα ORF112 - supporting the idea that this represents a support contact (**Figure 3.10B**). We observed that R258A Zα ORF112 binding to $T(CG)_6$ has less unfavorable entropy component than WT Zα ORF112 (-TΔS=11.1 kJ/mol versus 22.2 kJ/mol, **Table A.3** in **Appendix**). It indicates a potential role of R258 in a reduction of degrees of freedom of protein flexibility.

(A) S260L Zα ORF112 with $T(CG)_6$ **(B)** R258A Zα ORF112 with $T(CG)_6$

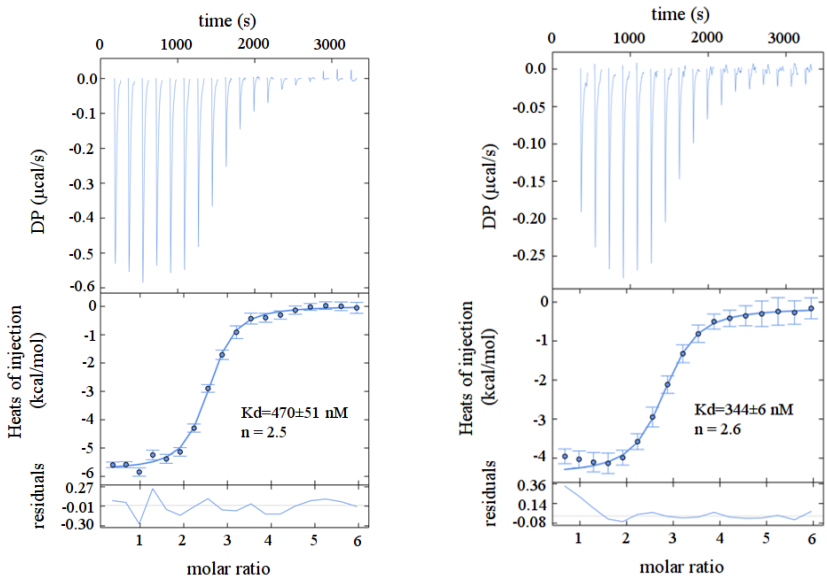


Figure 3.10: Isothermal titration calorimetry analysis of Z-DNA binding by Zα ORF112 mutants. Binding affinities of the Zα ORF112 mutants against $T(CG)_6$ were estimated from isothermal titration calorimetry data using a one-site model. **(A)** Titration of S260L Zα ORF112 against $T(CG)_6$ (1200 µM protein against 40 µM oligo) **(B)** Heat exchanged during R258A Zα ORF112 binding to $T(CG)_6$ (600 µM protein against 20 µM oligo).

3.3.4 Localization of ORF112 in the context of viral infection

In the previous paragraphs, we have demonstrated that Zα ORF112 shares many features with other domains of this family, both on the structural and biochemical level. We also observed some unique features in the interaction of this viral Zα domain with DNA. To see to which extent its unique features differentiate this domain on the functional level from other Zα domains, we characterized its localization in the context of viral infection in the CCB cells (common carp brain cell line). We also looked at the expression kinetics of the full-length ORF112 whether they are in agreement with a role in preventing immune responses.

One of the pending questions was whether ORF112 is expressed as the

annotated 287 amino acids protein (with the low complexity N-terminus) or as a shorter version starting at one of the internal methionine residues (M163, M187 or M222, accession number: BAF48926.1) consisting of almost exclusively of the Zα domain. Mass spectrometry analysis (2D-LC MS/MS) of the purified CyHV3 virions identified peptides encompassing the N-terminal part of ORF112, arguing that the protein is indeed expressed as annotated form. The function of the low-complexity part of ORF112 (the high occurrence of glutamines) is not known and requires further investigation, however, it is worth mentioning that such low complexity sequences are known to play an important role in RNP formations as stress granules. Having clarified what is the expressed form of the protein, we followed its localization in the context of infection (**Figure 3.11**). CCB cells were infected with CyHV3 and fixed, processed and visualized by the immunofluorescence microscopy at three-time points post-infection. Based on the polyclonal antibody staining, ORF112 was present both in the cytoplasm and nucleus of the infected cells. We observed an enrichment of ORF112 in the several regions of the cell membrane and inside the nucleoli. Additionally, ORF112 formed granular structures in the proximity and inside nucleus. These assemblies did not colocalize with CyHV3 structural proteins indicating that ORF112 is not enriched in the virion assembly/accumulation sites (**Figure 3.11**), in agreement with a non-structural role.

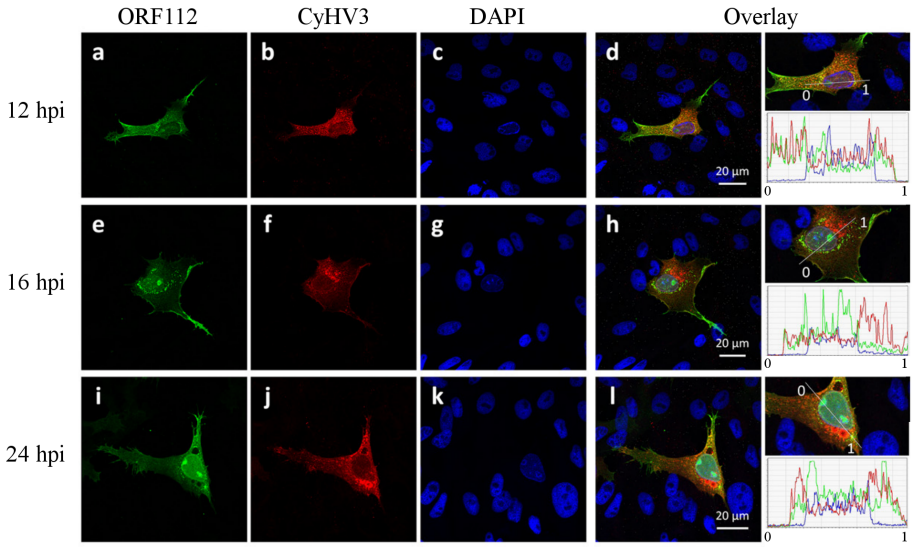


Figure 3.11: ORF112 subcellular localization during CyHV3 infection. CCB cells were infected with CyHV3 (0.01 plaque forming units per cell, PFU). At the specified times post-infection, cells were fixed and visualized for ORF112 (green signal; panels a, e, and i), CyHV3 structural proteins (red signal; panels b, f, and j), and DNA (panels c, g, and k). Composite images of the three stainings are shown in panels d, h, and l. The right side of these panels shows a magnification of the overlays and includes intensities quantifications of the three fluorochromes evaluated along the indicated line.

Previous studies have shown that Zαs of ADAR1 or DAI can be targeted to stress granules (SGs) ((Ng *et al.*, 2013) and Gabriel unpublished). However, the ORF112 granules signal did not colocalize with the immunofluorescence of HuR (the marker for SG). Apparently, CyHV3 has a mechanism to avoid stress granule formation (possibly ORF112 play a role in this process) during the infection. Nevertheless, we decided to induce the formation of stress granules by oxidative stress (arsenite treatment) in the infected cells to evaluate whether in these conditions ORF112 would accumulate in stress granules. Indeed, the arsenite treatment induced stress granules and ORF112 localized to these structures (**Figure 3.12**).

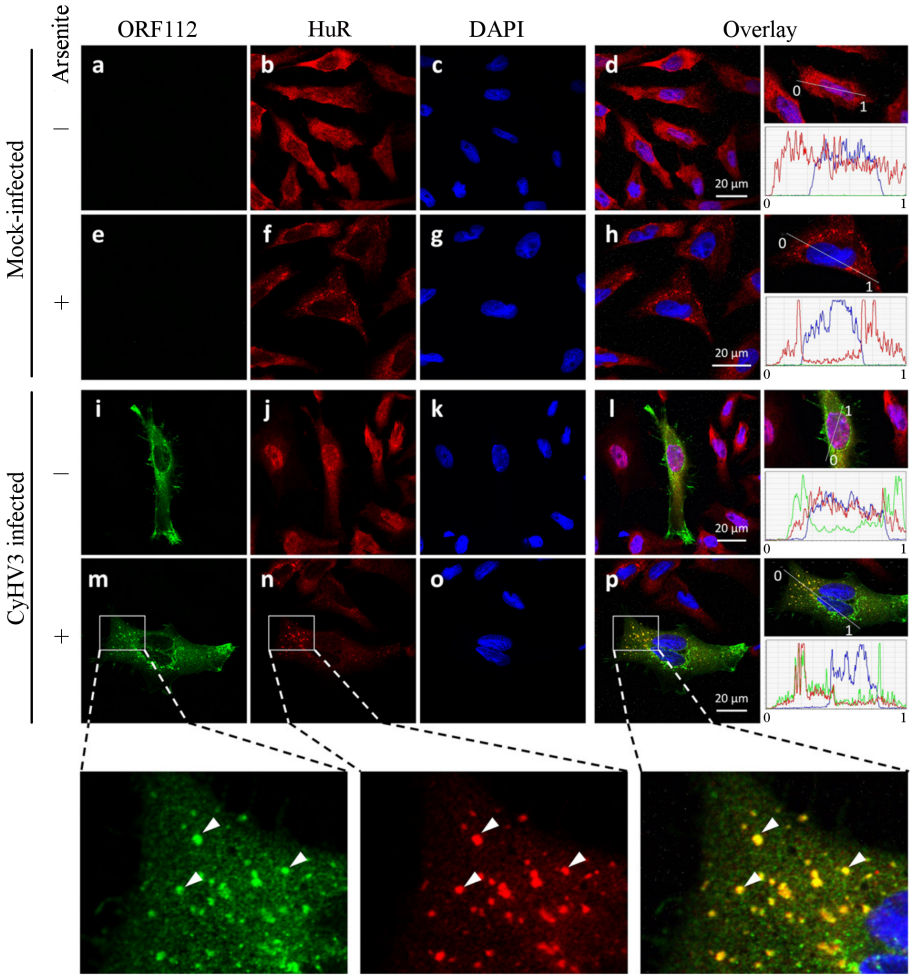


Figure 3.12: ORF112 localizes to stress granules during oxidative stress. CCB cells were infected with mock or CyHV3 (0.01 PFU). Cells were grown for 16 h and then incubated either with mock or 1 mM arsenite solution. Cells were then subjected to staining for CyHV3 ORF112 (green signal; panels a, e, i, and m), carp HuR/ELAVL1 (red signal; panels b, f, j, and n), and DNA (panels c, g, k, and o). Overlays of the three stainings are shown (panels d, h, l, and p). The right column of all panels illustrates the relative quantification of the intensities of the three fluorochromes assessed along the line indicated on the magnification of the overlay. The bottom row of the panels represents the magnification of a specified area of the panels m, n, and p. Arrows highlight some of the granules with overlapping signals of CyHV3 ORF112 and HuR. The abbreviation hpi stands for hours post-infection.

3.3.5 Potential Z-forming segments in the genome of cyprinid herpesvirus 3

The productive replication of viruses of *Herpesviridae* family, including cyprinid herpesvirus 3, requires that the capsid encapsulated DNA is released into the nucleus (nuclear replication). Nevertheless, the viral dsDNA or replication intermediates could be erroneously liberated in the cytoplasm. Alternatively, the capsid may be sensed and targeted to the proteasomal degradation exposing DNA to the cytosolic milieu. As cyprinid herpesvirus 3 infects fish which encode PKZ, Zα ORF112 might be a competitive inhibitor of PKZ. In agreement, ORF112 has been classified as immediate-early (IE) gene where the transcript has been observed around 2 h post-infection (hpi) (*in vitro*) and may function in early stages of infection (Ilouze *et al.*, 2012). In our assays (previous section), the ORF112 protein was present 12 hpi (although we did not evaluate earlier stages). We asked whether the genome of cyprinid herpesvirus 3 contains sequences with potential to form Z-DNA. We used custom scripts to assess available complete CyHV3 genome (accession: NC_009127.1, 295 146 bp) for the purine-pyrimidine repeats. We found 78 potential Z-forming sequences of at least 10 bp, randomly distributed throughout the genome (**Figure 3.13**).

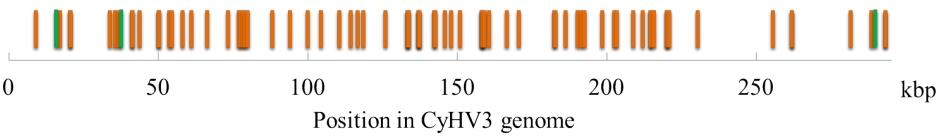


Figure 3.13: Distribution of potential Z-forming sequences in the genome of cyprinid herpesvirus 3. Starting positions of 78 sequences with Z-forming propensity (at least 10 bp) are shown along the viral genome as vertical lines. Three long perfect TG repeats (92, 74, 92 bp) are depicted in green.

Interestingly, three sequences were perfect, long GT repeats (92, 74, 92 bp) whereas in the shuffled (1000 times) viral genome 26 bp was the longest sequence with Z-forming propensity (average 17 bp with the standard deviation of 2). Two long GT repeats (92, 92 bp) are located close to

the ends of the genome. It is accepted that the longer purine-pyrimidine repeat the higher the probability of the B-to-Z transition (Pohl and Jovin, 1972; Peck and Wang, 1983). Hence, these long repeats might constitute putative sequences recognized by PKZ.

As we detected long perfect GT repeats in the genome of cyprinid herpesvirus, we decided to access the ability of Zα ORF112 to interact with $T(TG)_8$ or $T(CG)_8$ oligonucleotides and compare it with the prototypical Zα ADAR1. We tested both Zα domains in the gel shift mobility assay (**Figure 3.14**).

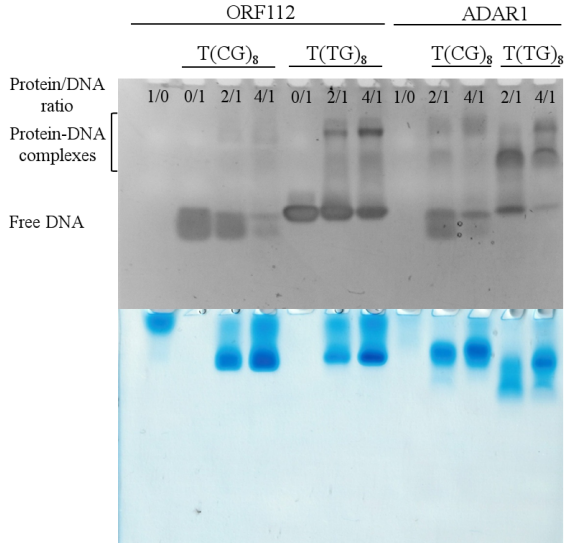


Figure 3.14: Complexes of Zα ORF112 with $T(TG)_8$ are different than Zα ADAR1·$T(TG)_8$. Mixtures of proteins and $T(CG)_8$ or $T(TG)_8$ were separated by the native gel and stained for nucleic acids (RedSafe, top) or protein (Coomassie blue, bottom). Zα ORF112 forms one major complex with $T(TG)_8$ whereas Zα ADAR1 seems to create two distinct protein-DNA complexes.

Both Zα ORF112 and ADAR1 shift similar amounts of $T(CG)_8$ (as judged by the amounts of the free oligonucleotide). The interaction of Zα ORF112 with $T(TG)_8$ shows a different pattern than that of Zα ADAR1. Although less $T(TG)_8$ oligonucleotide is shifted for ORF112, we observe a single complex (lane 6, 7), whereas Zα ADAR1 seems to form two distinct complexes (lane 11, 12). Lower bands (lane 11, 12) could reflect fewer Zα ADAR1 molecules bound to the nucleic acid and would be

consistent with the assumption that not entire oligonucleotide adopted the Z-conformation. Therefore, Zα ORF112 may be a better stabilizer of GT repeats in Z-conformation than Zα ADAR1 (at least under the conditions of our assay). One interesting observation is that the dye used to stain nucleic acids (RedSafe) exhibits differences in the staining potential of $T(CG)_8$ and $T(TG)_8$ bound to Zα domains. Thus, these repeats in the left-handed conformation differ in the propensity to intercalate this dye.

3.4 Discussion

The structure of the free Zα ORF112 revealed a domain-swapped dimer. We were intrigued by these results and carried out an extensive screening of different conditions and ligands to obtain crystals of the complex with Z-DNA. In this chapter, we discussed a structure of Zα ORF112 bound to $T(CG)_9$ DNA duplex. Based on the previous structure of apoprotein, we were not surprised to observe that Zα ORF112 recognizes the left-handed DNA conformation in a similar manner to other Zα domains (ADAR1, DAI, PKZ and E3L). The core of Z-DNA recognition is composed of three residues (Y257, N253, W274 - numbering for Zα ORF112) and their side chains adopt almost identical conformation among all Zα domains interacting with nucleic acids. Other viruses of the cyprinid herpesviral family have ORF112 and these residues are conserved among them suggesting that these proteins also possess functional Zα domains.

Despite many similarities with other Zα domain structures, Zα ORF112·$T(CG)_9$ complex revealed interesting features primarily due to the presence of the long DNA molecule. In previous studies (except for Zα PKZ) the minimal substrate ($T(CG)_3$) was utilized to analyze the details of the interaction with Z-DNA/Z-RNA. Here, we used 18 bp CG repeat which allowed uncovering interactions of Zα ORF112 extending beyond the minimal substrate. One of the key observations concerns the wing region which is oriented away from Z-DNA and, as a result, only P272

interacts with the nucleic acid. This unusual conformation of the wing region is stabilized by the protein-protein interactions and we propose that this on-DNA dimer enhances the stability of the complex formed by this viral protein. Mutagenesis experiments with Ser260, one of the residues mediating the protein-protein contacts supports our hypothesis. Therefore, Zα ORF112 may use the on-DNA dimerization as a mean to lock the Z-conformation decreasing the Z-to-B or Z-to-A transitions. Other proteins of this family may achieve such a stabilization using two covalently linked Zα domains. In fact, viral proteins with Zα contain one domain, whereas the metazoan proteins harbor two or more repetitions of Zα. An interesting exception within the viral Zα bearing proteins is CyHV1 ORF112 which contains two Zα domains connected by a very short linker of 2-3 amino acids. Such a short proximity between the two Zα domains would rather not allow interacting with the same Z-DNA/Z-RNA molecule. Yet, further studies are required to clarify how the tandem Zαs of CyHV1 ORF112 recognize the Z-conformation and what composes the functional unit interacting with the left-handed nucleic acids.

The recognition of nucleic acids vastly expands the number of pathogens detected but also imposes a complication in the discrimination between non-self and self. Cells evolved a range of nucleic acids sensors which detect and respond to the foreign genetic material. Different strategies are devised to distinguish own and pathogen nucleic acids including the absence host modifications in the foreign nucleic acids, sequence, structure, localization. The existence of Zα domains (in some host or viral proteins) suggests that even the transient conformations could be utilized to trigger the anti-viral responses. In fact, both E3L (Assarsson *et al.*, 2008) and ORF112 (Ilouze *et al.*, 2012) belong to the immediate–early viral genes class and their products are present since the initial stages of the infection which corroborate the idea that viral Zα interacts with its own nucleic acids, inhibiting recognition by cellular machinery. Interestingly, we identified three long, perfect GT repeats in the genome of CyHV3, which have a high potential to form Z-

conformation and could be one of the sites that require protection. Albeit the role of such repeats is not known, based on their terminal location, one may speculate about their involvement in the formation of the episome.

Our infection experiments demonstrate that CyHV3 does not induce the formation of stress granules allowing the translation of viral proteins. The contribution of the Z-DNA/Z-RNA binding to the inhibition of the stress granules formation could be studied by generation of the Zα ORF112 mutants and these experiments are in progress. In addition, ORF112 forms granule-like structures in the cytoplasm which do not colocalize with foci of the structural virion proteins. The role of this aggregation was not explored but might be implicated in the sequestration of the important stress granules components. The formation of the cytoplasmic ORF112 granule structures could be mediated by the N-terminal part of ORF112 containing the low-complexity region enriched in glutamines as multiple Gln/Asn-rich proteins are prone to form aggregates (Michelitsch and Weissman, 2000; Osherovich and Weissman, 2001). On the other hand, the induction of the oxidative stress (by arsenite) in the context of CyHV3 infections proves that ORF112 becomes enriched in stress granules just like the other Zα domains (Ng *et al.*, 2013). Additionally, previous studies and our observations indicate that the localization to stress granules requires Z-DNA/Z-RNA binding activity and is independent of the infection status. It suggests that these domains can also interact with certain host nucleic acids. As stress granules are mRNA repositories, it is likely that Zα domains could recognize the left-handed RNA in such structures. Yet, we do not know the origin, source of these RNAs. The next chapter describes our efforts undertaken to reveal the identity of sequences bound to the Zα domains using the domains of DAI as a probe.

In the previous chapter, we discussed the structure of the free Zα ORF112 which crystallized as a domain-swapped dimer. The gel filtration experiments proved that the dimer formation is dependent on the presence of the sulfate ions in solution but the predominant form is monomeric. The model of Zα ORF112 in complex with Z-DNA proved

that the monomer is the competent form to bind Z-DNA. The domain-swapping observed in the free protein structure is a crystallization artifact but provides interesting insights. First, the wing region which participates in the domain-swapping is rather flexible and under the favorable conditions (i.e. presence of sulfates) may be stabilized by the interactions with C-terminus from another monomer. In the Zα ORF112·Z-DNA complex, the wing region (especially β2 strand) is involved in the protein-protein interactions. Therefore, it is possible that the Z-DNA helix may participate in the determination of the location of the wing region.

Up to date, the only observed direct base contact for Zα domains was formed between conserved tyrosine residue and guanosine in syn conformation - the characteristic CH-π bond. Although it is an explicit interaction between a base and a residue it is not base-specific recognition as other residues found in syn conformation may be engaged in a similar interaction as demonstrated by the structures of Zα with other purine-pyrimidine repeats (Ha *et al.*, 2009). Unexpectedly, the structure of Zα ORF112 uncovered a novel direct base read-out. The guanosine interacting with Y257 (CH-π bond) of one of the monomers, forms two hydrogen bonds with R258 of the other chain. It is conceivable that other Zα domains also engage in the equivalent interactions if bound to longer substrates. As demonstrated by the current structure and Zα PKZ with $T(CG)_6$ (de Rosa *et al.*, 2013), the prototypical $T(CG)_3$ substrate used in all other studies it is not sufficient to capture all the interactions. Moreover, a short six bp CG repeat tends to have small distortion of the Z-DNA helix at the ends which can also affect the recognition by the Zα domain.

Acknowledgements

I would like to thank my supervisor, Alekos Athanasiadis, for the help and guidance he has provided during this project. I would like to acknowledge help from Alain Vanderplasschen's laboratory, which contributed data on ORF112 cellular localization. Finally, I am grateful Ewa Chrostek for careful reading and discussing this chapter.

Bibliography

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, 66(Pt 2):213–21. /cited on p. 115/

Assarsson, E., Greenbaum, J. A., Sundström, M., Schaffer, L., Hammond, J. A., Pasquetto, V., Oseroff, C., Hendrickson, R. C., Lefkowitz, E. J., Tscharke, D. C., Sidney, J., Grey, H. M., Head, S. R., Peters, B., and Sette, A. (2008). Kinetic analysis of a complete poxvirus transcriptome reveals an immediate-early class of genes. *Proc. Natl. Acad. Sci. U. S. A.*, 105(6):2140–5. /cited on p. 141/

Ban, C., Ramakrishnan, B., and Sundaralingam, M. (1996). Crystal structure of the self-complementary 5'-purine start decamer d(GCGCGCGCGC) in the Z-DNA conformation. I. *Biophys. J.*, 71(3):1215–21. /cited on p. 116/

Brennan, R. G., Westhof, E., and Sundaralingam, M. (1986). Structure of a Z-DNA with two different backbone chain conformations. Stabilization of the decadeoxyoligonucleotide d(CGTACGTACG) by [Co(NH3)6]3+ binding to the guanine. *J. Biomol. Struct. Dyn.*, 3(4):649–65. /cited on p. 116/

Cole, J. L. (2007). Activation of PKR: an open and shut case? *Trends Biochem. Sci.*, 32(2):57–62. /cited on p. 112/

Costes, B., Fournier, G., Michel, B., Delforge, C., Raj, V. S., Dewals, B., Gillet, L., Drion, P., Body, A., Schynts, F., Lieffrig, F., and Vanderplasschen, A. (2008). Cloning of the koi herpesvirus genome as an infectious bacterial artificial chromosome demonstrates that disruption of the thymidine kinase locus induces partial attenuation in Cyprinus carpio koi. *J. Virol.*, 82(10):4955–64. /cited on p. 117/

de Rosa, M., Zacarias, S., and Athanasiadis, A. (2013). Structural basis for Z-DNA binding and stabilization by the zebrafish Z-DNA dependent protein kinase PKZ. *Nucleic Acids Res.*, 41(21):9924–33. /cited on p. 113, 116, 118, 122, 124, 143/

Delano, W. . (2002). The PyMOL Molecular Graphics System. /cited on p. 116/

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.*, 36(Web Server issue):W465–9. /cited on p. 116/

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, 66(Pt 4):486–501. /cited on p. 115/

Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, 52(5):696–704. /cited on p. 116/

Ha, S. C., Choi, J., Hwang, H.-Y., Rich, A., Kim, Y.-G., and Kim, K. K. (2009). The structures of non-CG-repeat Z-DNAs co-crystallized with the Z-DNA-binding domain, hZ alpha(ADAR1). *Nucleic Acids Res.*, 37(2):629–37. /cited on p. 143/

Ha, S. C., Lokanath, N. K., Van Quyen, D., Wu, C. A., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K. K. (2004). A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 101(40):14367–72. /cited on p. 113/

Houtman, J. C., Brown, P. H., Bowden, B., Yamaguchi, H., Appella, E., Samelson, L. E., and Schuck, P. (2007). Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. *Protein Sci.*, 16(1):30–42. /cited on p. 118/

Ilouze, M., Dishon, A., and Kotler, M. (2012). Coordinated and sequential transcription of the cyprinid herpesvirus-3 annotated genes. *Virus Res.*, 169(1):98–106. /cited on p. 138, 141/

Kabsch, W. (2010). XDS. *Acta Crystallogr. D. Biol. Crystallogr.*, 66(Pt 2):125–32. /cited on p. 115/

Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A., and Schuck, P. (2012). High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.*, 84(11):5066–73. /cited on p. 118/

Kim, D., Hur, J., Park, K., Bae, S., Shin, D., Ha, S. C., Hwang, H.-Y., Hohng, S., Lee, J.-H., Lee, S., Kim, Y.-G., and Kim, K. K. (2014). Distinct Z-DNA binding mode of a PKR-like protein kinase containing a Z-DNA binding domain (PKZ). *Nucleic Acids Res.*, 42(9):5937–48. /cited on p. 113/

Kim, Y.-G., Muralinath, M., Brandt, T., Pearcy, M., Hauns, K., Lowenhaupt, K., Jacobs, B. L., and Rich, A. (2003). A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, 100(12):6974–9. /cited on p. 113/

Krissinel, E. and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, 372(3):774–97. /cited on p. 116/

Le, V. H., Buscaglia, R., Chaires, J. B., and Lewis, E. A. (2013). Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal. Biochem.*, 434(2):233–41. /cited on p. 118/

Michel, B., Fournier, G., Lieffrig, F., Costes, B., and Vanderplasschen, A. (2010). Cyprinid Herpesvirus 3. *Emerg. Infect. Dis.*, 16(12):1835–1843. /cited on p. 117/

Michelitsch, M. D. and Weissman, J. S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. U. S. A.*, 97(22):11910–5. /cited on p. 142/

Neukirch, M., Böttcher, K., and Bunnajirakul, S. (1999). Isolation of a virus from koi with altered gills. *Bull. Eur. Assoc. Fish Pathol.*, 19(5):221–224. /cited on p. 117/

Ng, S. K., Weissbach, R., Ronson, G. E., and Scadden, A. D. J. (2013). Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules. *Nucleic Acids Res.*, 41(21):9786–99. /cited on p. 114, 136, 142/

Osherovich, L. Z. and Weissman, J. S. (2001). Multiple Gln/Asn-Rich Prion Domains Confer Susceptibility to Induction of the Yeast [PSI+] Prion. *Cell*, 106(2):183–194. /cited on p. 142/

Peck, L. J. and Wang, J. C. (1983). Energetics of B-to-Z transition in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 80(20):6206–10. /cited on p. 139/

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–a

visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25(13):1605–12. /cited on p. 116/

Pohl, F. M. and Jovin, T. M. (1972). Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.*, 67(3):375–396. /cited on p. 139/

Rothenburg, S., Deigendesch, N., Dittmar, K., Koch-Nolte, F., Haag, F., Lowenhaupt, K., and Rich, A. (2005). A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl. Acad. Sci. U. S. A.*, 102(5):1602–7. /cited on p. 113/

4

In vivo interactions of Zα domains

Abstract

The Z-DNA/Z-RNA binding domains (Zα) interact with nucleic acids in the left-handed conformation. Evidence shows that Zα domains are enriched in stress granules (SGs), cytoplasmic non-membrane structures composed of stalled pre-initiation ribosomal complexes. Additionally, the Zα localization to SGs requires their Z-DNA/Z-RNA binding activity suggesting that these domains likely interact with Z-RNA in these structures. Yet, we have only a limited knowledge about the origin of nucleic acids bound to these domains. To fill the knowledge gap concerning *in vivo* Zα interactors, we created cell lines expressing Z$\alpha\beta$ domains from the DNA-dependent activator of IFN-regulatory factors (DAI) as well as its mutant incompetent of the Z-DNA/Z-RNA binding. To answer the question of the nature of nucleic acids bound to Z-DNA/Z-RNA domains under stress conditions we developed the tail-CLIP protocol involving UV cross-linking of protein-RNA complexes followed by immunoprecipitation, library construction and high-throughput sequencing. Although we could not attribute any specific sequences bound by the wild-type Z$\alpha\beta$, our results provide valuable information for future efforts.

Publication

Unpublished results authored by Luisa Gabriel, Krzysztof Kuś and Alekos Athanasiadis.

Contribution

Alekos Athanasiadis initiated and supervised this study. Alekos Athanasiadis, Luisa Gabriel and Krzysztof Kuś designed experiments. Krzysztof Kuś developed the tail-CLIP method, analyzed preliminary sequencing results. Luisa Gabriel established stable cell lines, evaluated

the localization of Z$\alpha\beta$ domains and assisted in the development of tail-CLIP protocol.

4.1 Introduction

The Z-DNA/Z-RNA binding (Zα) domains can bind and stabilize left-handed nucleic acid helices. The Zα domains have been extensively studied by biophysical/structural methods and such studies revealed that Zα domains recognize the shape of the left-handed helices (both RNA and DNA) (Schwartz, 1999; Schwartz *et al.*, 2001; Ha *et al.*, 2004; Athanasiadis *et al.*, 2005; de Rosa *et al.*, 2013; Ha *et al.*, 2008). Nevertheless, our knowledge about the intracellular nucleic acid ligands of Zα domains is very limited. This may be explained by the fact that the Z-conformation is transient and it is technically challenging to capture real interaction complexes. Secondly, these domains are found only in proteins involved in the interferon responses and it has been assumed that binding to the nucleic acid occurs in the context of infection which increases the difficulty. The so far available information about *in vivo* Zα bound nucleic acids was obtained from pull-down experiments. In one of the studies, Zα ADAR1 (adenosine deaminase acting on dsRNA, compare **Chapter 1**) was used as a probe to detect Z-DNA in the human genome (Li *et al.*, 2009). First, cells were fixed with formaldehyde and then the purified probes (Zα or Zα DNA binding mutant) were incubated with the cells followed by the second cross-linking. Subsequently, cells were lysed and nuclei were collected and sonicated. The Zα-DNA complexes were immunoprecipitated with beads against Zα fused with Strep-tag. Then, the cross-linking was reversed and fragments were cloned into a vector and sequenced. This study found around 200 genomic Z-DNA hotspots, of which nearly fifty were located in the centromeres. The complex two-step cross-linking used in this study may have introduced artifacts and some of the detected regions underwent the B-Z transition due to the procedure deployed. In any case, the biological relevance of this study relates to areas of the genome that may adopt Z-conformation but not for

the function of Zα domains that normally reside in the cytoplasm. Thus in this study Zα domain was used more as a tool for the identification of Z-DNA segments rather than for clarifying the function of the domains.

In a second study, the evidence is presented for an interaction between ribosomal RNA (rRNA) and Zα domains. Here the authors focused on the Zα domain from the RNA editing enzyme ADAR1, which is expressed in two isoforms: constitutive (p110, only Zβ domain) and interferon-induced (p150, with both Zα and Zβ domains). It is shown that Zα ADAR1 binding to ribosome results in translation inhibition. On the contrary, an RNA-binding mutant Zα domain has almost no impact on the translational efficiency. Based on this result the authors proposed that the interferon-induced ADAR1 p150 containing Zα domain can restrict viral replication through translation inhibition (Feng *et al.*, 2011). However, more recent studies have classified ADAR1 (both p110 and 150 forms) as a general pro-viral protein. Indeed, it is demonstrated that the major ADAR1 function in innate immunity is the suppression of interferon responses (Mannion *et al.*, 2014; Liddicoat *et al.*, 2015). Further studies are needed to clarify the impact of Zα ADAR1 on the inhibition of viral replication.

It has been observed that under stress conditions (i.e. oxidative stress) DAI, ADAR1 (p150 isoform) and ORF112 (**Chapter 3**) are found in stress granules (SG) (Deigendesch *et al.*, 2006; Ng *et al.*, 2013; Kuś *et al.*, 2015). Stress granules are ribonucleoparticles (RNPs) formed by stalled ribosomes and RNA binding proteins during cellular translation arrest. These structures may be important for the control of cell survival. When stress conditions cease, mRNA trapped in SGs may become again available for translation (Anderson and Kedersha, 2008) or be degraded in the processing-bodies (PBs), which are dynamically linked to SGs (Kedersha *et al.*, 2005). It has been demonstrated that Zα domains are required for SGs association of these proteins. Moreover, it has been established that artificial fusion proteins with Zα domains are also targeted to stress granules. Importantly, mutations of the residues involved in Z-DNA/Z-RNA binding abolish localization to stress granules. As SGs are cytoplasmic RNPs, the most likely partner for Zα domains is

Z-RNA (Ng *et al.*, 2013) which is consistent with the report discussed above. These observations have important implications as they suggest that infection is not the only condition under which these domains are functional. It poses a question whether rRNA is a sole Zα binding partner and, if not, what is the identity of the other potential nucleic acid substrates.

DAI was demonstrated to induce interferon responses upon detection of nucleic acids through NF-κB or IRF3/TBK1 pathways (Takaoka *et al.*, 2007). DAI contains two functional Z-DNA/Z-RNA binding domains (Z$\alpha\beta$) and a suggested B-DNA binding (D3) region (**Figure 1.11**) and due to alternative splicing there are several isoforms of the protein (Rothenburg *et al.*, 2002). Although the D3 region was essential for B-DNA binding *in vitro*, Z$\alpha\beta$ domains were required for an efficient activation of IFNβ production in these experiments (Wang *et al.*, 2008). The C-terminal DAI region contains RHIM motifs which mediate interactions with RIP1 and RIP3 (activators of NF-κB pathway) (Kaiser *et al.*, 2008). A more recent study implicated DAI in the virus-mediated programmed necrosis (necroptosis), triggered by interaction with RIP3 (via RHIM motifs) (Upton *et al.*, 2012).

Nevertheless, not all cell types rely on DAI to activate interferon responses. Several studies indicated that there is a redundancy of cytoplasmic DNA sensors and the depletion of DAI in some cell types does not alter interferon production (Wang *et al.*, 2008). On the other hand, DAI may control viral titers independently of the interferon production. The study of herpes simplex virus 1 (HSV1) infection in HepG2 cells revealed that DAI has an impact on the viral replication through interaction with ICP0 (a protein with E3 ubiquitin ligase activity, modifying and directing proteins for proteasomal degradation) (Pham *et al.*, 2013).

In summary, we lack a comprehensive analysis of nucleic acids bound to Zα domains. Here, we studied the cellular localization of Zα DAI domains (or its mutant) using stable cell lines expressing these domains as GFP fusion proteins. We then use these cell lines for the isolation of

nucleic acids Zα partners. We have confirmed that Z$\alpha\beta$ DAI (but not its Z-DNA/Z-RNA binding mutant) is enriched in the stress granules. We also describe the development of the tail-CLIP methodology designed to uncover sources of nucleic acids bound to Zα domains during stress conditions.

The tail-CLIP method is based on UV cross-linking of proteins and RNA followed by immunoprecipitation and sequencing. We have obtained preliminary results of the next generation sequencing of RNAs bound to Z$\alpha\beta$ DAI or its mutant. However, as we did not recover specific sequences associated with Z$\alpha\beta$ DAI we discuss possible future improvements of this method.

4.2 Materials and Methods

Plasmid construction

The insert containing N-terminal GFP and two Z-DNA binding domains of human DAI (ZBP1/DLM-1, residues 2-165 of NP_110403.2) was cloned into a pBABE vector (referred as Z$\alpha\beta$). The linker between GFP and two Zαs has a Tev cleavage site (used to release the Z-DNA binding domains from the beads) and an S-tag sequence. Mut Z$\alpha\beta$ refers to quadruple Z$\alpha\beta$ mutant - N46A, Y50A, N141A and Y145A. Mut Z$\alpha\beta$ is not capable of Z-DNA/Z-RNA binding. Plasmids were used to establish stable A549 cell lines.

Stable cell line establishment, arsenite and UV treatment

A549 (adenocarcinomic human alveolar basal epithelial) cells were cultured in high glucose DMEM medium supplemented with 10% fetal bovine serum, L-glutamine and penicillin/streptomycin at 37 °C, 5% CO_2. A549 stable cell lines expressing: Z$\alpha\beta$, Mut Z$\alpha\beta$ were generated using retroviral expression vector pBABE. First, the infectious viral-like particles were produced in HEK293-G cell line (plated onto a 10 cm dish) by co-transfection with 5 µg of pBABE and 2 µg of pVSV-G (helper plasmid

encoding packing protein) in the presence of Lipofectamine LTX and Plus Reagent (ThermoFisher). Infected cells were grown for 3 days with the medium changed every day. In the last day the medium was harvested, filtered through 0.45 µm filter and frozen at -80 °C. Infections of A549 cell lines were performed in a 6-well dish at 30-40% confluency. Infection was performed with two doses of virus 750 (250 medium) or 250 (750 medium) µl, in the presence of 8 µg/ml of polybrene. After 24 h, the medium was replaced. Cells were grown to confluency, trypsinized and seeded onto a new plate containing medium with puromycin (4 µg/ml) for selection. Arsenite treatment was used to induce stress granules formation by oxidative stress. Briefly, nearly confluent plates with A549 cell lines were washed twice with PBS. Each 15 cm plate was supplied with 15 ml of DMEM medium with 0.5 mM sodium arsenite. The plates were incubated at 37 °C, 5% CO_2 for 30 min. Then, the medium was discarded, plates were washed twice with PBS, and a new complete medium (15 ml) was provided. Then, plates were kept for 30 min in the incubator. The medium was removed and plates were washed twice with PBS, 15 ml of PBS was added and cells were UV treated (1 mJ, UV device: Stratalinker). Cell pellets were stored at -80 °C until use. Each cell line was evaluated for the presence of the correct insert by PCR and Sanger sequencing.

Tail-Clip

Approximately 1.5 g of frozen cell pellet of arsenite-treated, UV cross-linked A549 cell line (either Z$\alpha\beta$ or Mut Z$\alpha\beta$) was placed on ice for 20 min. Cells were disrupted with 3 ml of lysis buffer (10 mM TRIS pH 7.4, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40) supplemented with protease inhibitors (complete, Mini, EDTA-free - Roche, following manufacturer recommendation) and 5 µl of Turbo DNAse (Thermo Fisher Scientific, 2 U/µl). Cell lysates were kept on ice for 1 h with vigorous pipetting every 10 min. After lysis, cell extracts were supplemented with additional 5 µl of Turbo DNAse and 3 µl of RNase I (Ambion, 100 U/µl). Each lysate was aliquoted in ice-cold 1.5 ml tubes (around 840 µl in each of the five tubes). Mixtures were incubated at 37 °C for 10 min with 1100

rpm rotation and then placed on ice for 5 min. The insoluble fraction was removed by centrifugation at 13000 rpm at 4 °C for 20 min. Meanwhile, 100 µl of anti-GFP beads (GFP-Traps, ChromoTek) was dissolved in 1 ml of dilution/wash buffer (10 mM TRIS pH 7.4, 150 mM NaCl, 0.5 mM EDTA, 0.05% Tween-20) and washed 3 times with this buffer. Beads were resuspended in 200 µl of wash/dilution buffer. After centrifugation, the supernatant was transferred to new tubes and centrifuged for additional 15 min. Pre-cleared lysate (either Z$\alpha\beta$ or Mut Z$\alpha\beta$) was diluted in 9 ml of dilution/wash buffer with protease inhibitors. The washed beads were added to each of the diluted samples and incubated at 4 °C for 1.5 h, rotating. Next, beads were separated (10 min at 4 °C) and the supernatant was removed. The beads were washed 3 times with 12 ml of ice-cold high-salt wash buffer (10 mM TRIS pH 7.4, 1 M NaCl, 0.5 mM EDTA, 0.05% Tween-20) and kept on ice for 5 min between washes (separation time 10 min). Then, beads were removed with 1 ml of high-salt wash buffer and transferred to a new Eppendorf tube. Beads were washed 3 times with 1 ml of PNK buffer (20 mM TRIS pH 7.4, 10 mM $MgCl_2$, 0.1 % Tween-20). Following removal of the last wash, beads were resuspended in 30 µl of 3' dephosphorylation mix: 0.75 µl T4 Polynucleotide Kinase (NEB, 10/ µl/), 1 µl anti-RNase (Ambion, 15 10/ µl) in 1x DeP buffer (final: 70 mM TRIS, pH 6.5, 10 mM $MgCl_2$, 1 mM dithiothreitol). Samples were incubated at 37 °C, 1100 rpm mixing for 20 min. Beads were separated and washed with 1 ml PNK buffer, high-salt wash buffer and again twice with PNK buffer. Wash solution was discarded and beads were resuspended in 30 µl of ligation solution (refer to **Figure 4.3A**): 1.5 µl T4 RNA ligase 1 (NEB, 10 U/µl), 2.25 µl pre-adenylated, biotinylated L3 adaptor (20 µM), 1 µl anti-RNase, 6 µl PEG400 (100%), 1x ligation buffer (50 mM TRIS pH 7.8, 10 mM $MgCl_2$, 1 mM dithiothreitol). Ligation mixtures were incubated overnight at 16 °C in thermomixer at 1100 rpm shaking. Next, 500 µl of PNK buffer was added to each sample. Then beads were washed 2 times with 1 ml of high-salt buffer and 3 times with PNK buffer. Protein-RNA complexes were cleaved of beads with 114 µl of AcTev mix: 1.5 µl AcTev (Novex, 10 U/µl), 2 µl anti-RNase, 1 mM DTT, 1x AcTev buffer (the component of the protease kit). Digestion was performed at 4 °C for 1

h, shaking at 1100 rpm. To facilitate complex recovery, 6 µl of 2 M NaCl was added to each sample (final concentration 100 mM NaCl). Samples were incubated at 4 °C, 1100 rpm for additional 10 min. Afterward, beads were separated and the supernatant containing released RNA-protein complexes was transferred to new tubes. To release RNA cross-linked to protein, 2 µl of proteinase K (NEB, 0.8 U/µl) was added and incubated 20 min shaking at 1100 rpm, 37 °C. In order to inhibit proteinase K samples were supplemented with PMSF (200 mM in isopropanol) to the final concentration of 10 mM and incubated at room temperature for 5 min. Meanwhile, 5 µl of MyOne Streptavidin T1 (T1 beads, Invitrogen) was washed twice with 200 µl of StrepBead Wash Buffer (100 mM Tris–HCl, pH 7, 1 M NaCl, 10 mM EDTA, 0.1% Tween-20). Beads were resuspended in 100 µl of StrepBead Wash Buffer. Next, proteinase K treated RNA samples were mixed with pre-washed T1 beads and incubated at 4 °C for 10 min. After incubation, beads were separated for 2 min and the supernatant was discarded. The beads were washed 3 times with 100 µl of MyOne Wash Buffer (100 mM Tris–HCl pH 7, 2 M NaCl, 10 mM EDTA, 0.2% Tween-20) and twice with 100 µl of NT2 buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 0.05 % NP-40) and finally with 200 µl of RNase-free water. The reverse transcription reaction was performed on washed beads with P3_sh primer (sequence in **Appendix**) and 0.5 µl Superscript III (Invitrogen, 200U/µl) incubating at 40 °C for 15 min and then 15 min at 50 °C (in PCR thermocycler). Samples were treated with 1 µl of RNase H (NEB, 5 U/µl) and 1 µl of RNase Cocktail (Invitrogen, RNase A 0.5 U/µl, RNase T1 20 U/µl) for 15 min at 37 °C. Beads were subsequently washed three times with 100 µl of MyOne Wash buffer (100 mM Tris–HCl pH 7, 2 M NaCl, 10 mM EDTA, 0.2% Tween-20), 3 times with 100 µl of NT2 buffer and once with 200 µl of H_2O. In order to provide a priming region for subsequent PCR, poly-A tailing of cDNA was carried out. 50 µl of tailing mixture was composed of 0.5 µl of terminal transferase (TdT, NEB 20 U/µl), 0.75 µl ATP (2.5 mM) and 1x TdT buffer and $CoCl_2$ (included in the kit). The reaction was incubated at 37 °C for 5 min. Washes with MyOne Wash buffer, NT2 buffer and water were executed as above. The cDNA was eluted twice in 13 µl of H_2O (boiling

beads at 95 °C for 3 min). 20 µl of eluted cDNA was used as a template for PCR amplification with Phusion Hot Start Flex polymerase (NEB). Primers used for Z$\alpha\beta$: MiSeq_F1, MiSeq_R1 and Mut Z$\alpha\beta$: MiSeq_F2, MiSeq_R2 (see **Appendix**)). Initial annealing was done at 45 °C for 4 cycles, and then at 55 °C for 37 cycles. The product of the first PCR was purified with Agencourt AMPure XP beads (Beckman Coulter) and used in the second PCR to introduce sequences complementary to the cell of MiSeq sequencing chip. Size distribution of sequences in the library was evaluated by BioAnalyzer. qPCR was used to determine the amount of sequences competent to anneal with the adaptor of the sequencing chip. The sequencing was performed with 1 µl of pooled 2 nM libraries on MiSeq instrument (Illumina) (pair-end, 250 cycles run).

4.3 Results

4.3.1 Search for the nucleic acids interacting with Zα domains

Z$\alpha\beta$ DAI localizes to cytoplasmic stress granules

Previous studies have shown that proteins with Zα domains are targeted to stress granules. Zα domains were required and sufficient to direct proteins to these stress-induced structures (Deigendesch *et al.*, 2006; Ng *et al.*, 2013; Kuś *et al.*, 2015). Additionally, localization to SGs was dependent on the Z-DNA/Z-RNA binding activity of these domains suggesting that the left-handed nucleic acids are present in the absence of infection. We intended to reveal the identity of nucleic acids bound to Zα domains. For this purpose, we created A549 stably transformed cell lines expressing either Z$\alpha\beta$ DAI or its mutant (Mut Z$\alpha\beta$) as GFP fusion proteins. Mut Z$\alpha\beta$ DAI has mutations in both domains of the key residues responsible for the interaction with Z-DNA/Z-RNA (N46A, Y50A, N141A, and Y145A). First, we evaluated the cellular distribution of our fusion proteins. Under standard conditions, both Z$\alpha\beta$ and Mut Z$\alpha\beta$ DAI showed a diffuse distribution in both the cytoplasm and nucleus, whereas, the SGs marker (TIAR; purple signal) was present only in the cytoplasm (**Figure 4.1**). During oxidative stress (induced by arsenite treatment), TIAR accumulated in the cytoplasmic stress granules (**Figure 4.1**). Only Z-DNA/Z-RNA binding competent Z$\alpha\beta$ but not Mut Z$\alpha\beta$ localizes to SGs (**Figure 4.1**, white arrows). We concluded that Z$\alpha\beta$, like other Zα domains, can be targeted to SGs and this localization is mediated by the binding to nucleic acids. As stress granules are cytoplasmic structures composed of stalled pre-initiation complexes, we assumed that the Z$\alpha\beta$ may be involved in interactions with Z-RNA. We then decided to use the iCLIP (individual-nucleotide resolution UV cross-linking and immunoprecipitation) methodology to capture and sequence RNAs bound to Z$\alpha\beta$. Mut Z$\alpha\beta$ was used as a

negative control as it does not bind nucleic acids *in vitro* and it does not show SG localization.

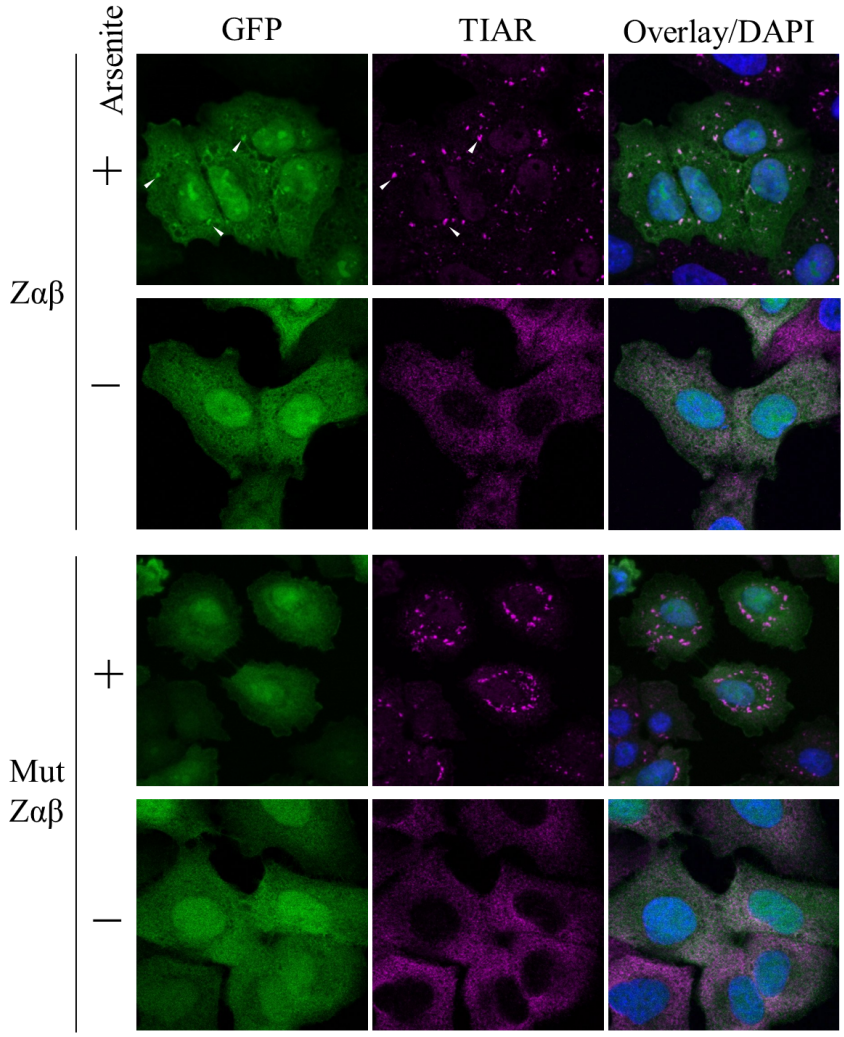


Figure 4.1: Z$\alpha\beta$ enrichment in stress granules is dependent on the residues important for the recognition of Z-DNA/Z-RNA. Z$\alpha\beta$ or Mut Z$\alpha\beta$ DAI GFP fusion expressing A549 cell lines were mock or arsenite-treated. The cells were permeabilized and stained for the stress granule marker TIAR (purple color). DAPI staining of the nucleus was included. The last panel presents an overlay of GFP, TIAR and DAPI channels. White arrows indicate representative stress granules with the enrichment in the Z$\alpha\beta$ DAI. Mut Z$\alpha\beta$ remained diffuse in arsenite-treated cells.

Development of the tail-CLIP protocol

In the first experiments with the objective to characterize RNAs interacting with Zα domains, we tried to employ iCLIP (Huppertz *et al.*, 2014) but we encountered problems with library preparation. Moreover, the iCLIP protocol involves many nucleic acid precipitations and it is time-consuming, which makes optimization difficult. Therefore, we directed our interest towards a modified version of iCLIP named FAST-iCLIP introducing a biotinylated adaptor facilitating purifications (Flynn *et al.*, 2015). Unfortunately, we could not amplify any sequences using FAST-iCLIP. Thus, we have introduced many alterations to the FAST-iCLIP method. One of the most significant changes concerns the omission of cDNA circularization step (a characteristic feature of iCLIP methodology). Instead, a second priming site for PCR is created by terminal transferase enzyme introducing a poly-A tail. We named the current protocol tail-CLIP. **Figure 4.2** summarizes our experimental approach. Briefly, we started by growing A549 stably expressing Z$\alpha\beta$ or Mut Z$\alpha\beta$ DAI fused N-terminally to GFP (compare **Materials and Methods**). Cells (~90% confluent) were treated with arsenite and subjected to UV cross-linking. Cells were collected and frozen at -80 °C until use. During cell lysis, we included partial RNA degradation (RNase I) to fragment potential high-molecular RNA complexes. Next, we performed immunoprecipitation with magnetic anti-GFP beads. In order to ligate L3 adaptor (5' pre-adenylated, 3' biotin residue), we dephosphorylated the 3' end of RNAs. In the next step, we released the RNA-protein complexes from anti-GFP beads with AcTev (cleaving off GFP). Subsequently, proteins were digested with proteinase K, leaving the RNA with a peptide at the crosslink site. Inactivation of proteinase K was achieved by treatment with phenylmethylsulfonyl fluoride (PMSF). Further, we used the streptavidin magnetic beads to capture and purify RNA and perform reverse transcription. After cDNA was produced, we digested the RNA strand with RNase H. To provide another priming site for PCR (one is provided by L3 adaptor), we carried out tailing reaction with terminal transferase (TdT adds poly-A to 3' end). Finally, cDNA was

eluted from the beads and used in PCR to prepare libraries.

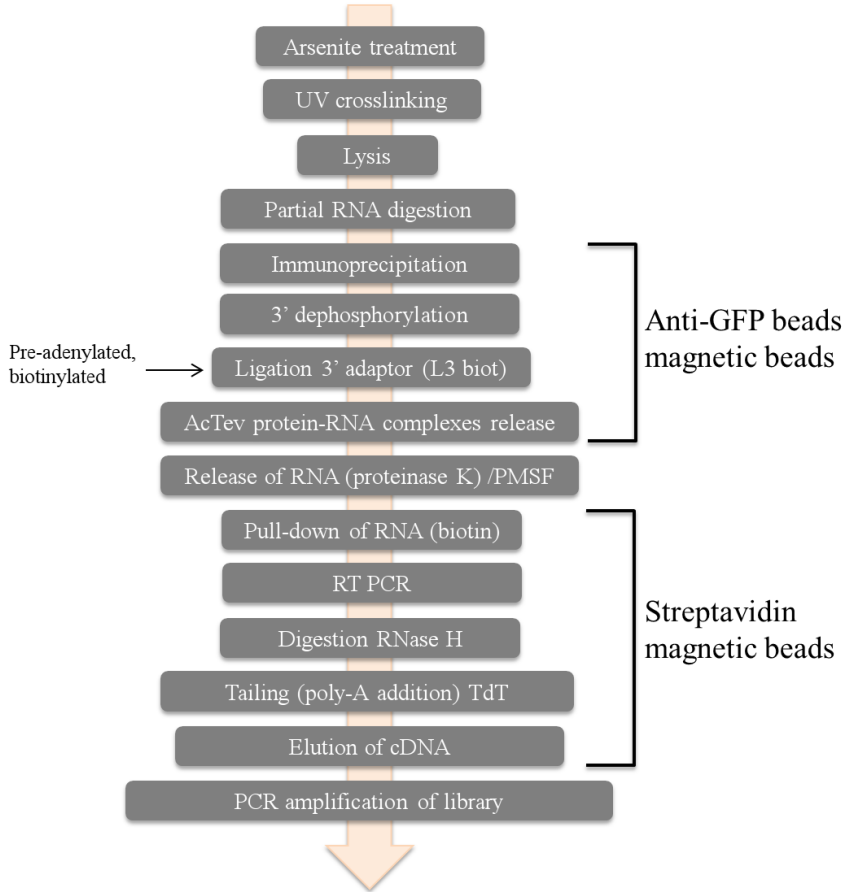
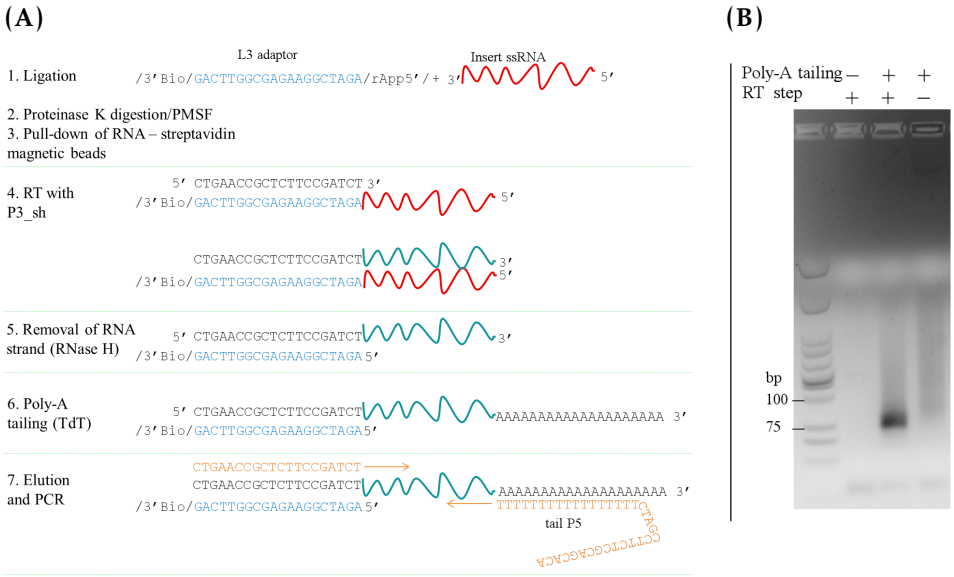


Figure 4.2: Schematic representation of tail-CLIP protocol. The protocol is based on UV RNA-protein cross-linking followed by immunoprecipitation and sequencing. Purifications require two types of magnetic beads: anti-GFP (to immunoprecipitate protein) and streptavidin (to purify and process RNA with ligated L3 adaptor). Preparation of the library is based on PCR amplification of the obtained cDNA. The ligated L3 adaptor and poly-A tail serve as priming sites.

To evaluate whether our protocol may be successful in amplification of cDNA from RNAs cross-linked to Zα domains, we decided to perform experiments with a single-stranded RNA oligonucleotide of known sequence (**Figure 4.3A**). For this purpose, we ligated L3 adaptor to a synthetic RNA oligonucleotide. We also included proteinase K and PMSF treatment. Ligation was performed in the absence of anti-GFP beads but the steps 2-7 of the protocol (see **Figure 4.3A**) were performed as depicted

in **Figure 4.2**.



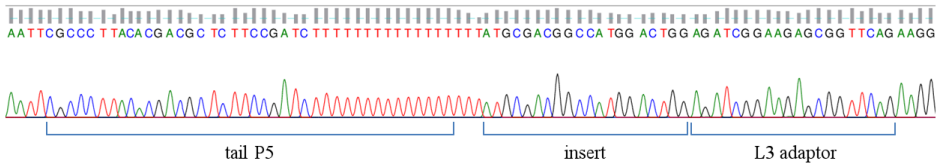


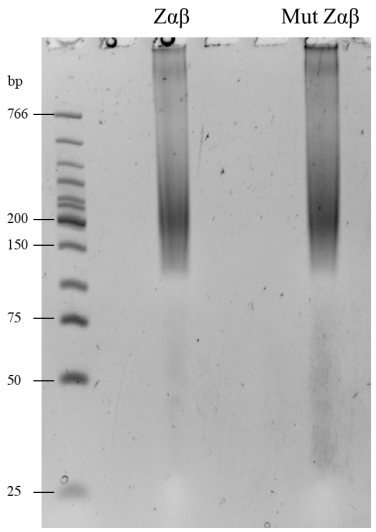
Figure 4.3: Assessment of the tail-CLIP methodology with an RNA oligonucleotide. (A) Schematic representation of the steps of the partial tail-CLIP protocol (no protein pull-down) with the emphasis on the primer arrangements. **(B)** Agarose gel separation of the amplified cDNA PCR products prepared following steps from (A). To assure the specificity of the amplification we either omitted the addition of terminal transferase (marked poly-A tailing) or reverse transcriptase (denoted as RT step). **(C)** The sequencing trace of the PCR product prepared by the partial tail-CLIP method. The fragment is composed of an expected insert (RNA, Gabra-3 gene) flanked by primers used for PCR.

Indeed, following all steps, we could amplify a PCR fragment of the expected size, whereas in the controls without terminal transferase (poly-A tail) or reverse transcriptase (RT) we have not observed specific product (**Figure 4.3B**). To assure that the amplified fragment consisted of all expected elements and was not a result of the unspecific amplification,

we excised a band from the gel and cloned it to the vector and sequenced. Sanger sequencing revealed that insert sequence (RNA) was encapsulated between L3 adaptor and tail P5 primer (compare **Figure 4.3A, C**).

Encouraged by the results from the experiments with synthetic RNA of known sequence we proceeded to evaluate the complete protocol (**Figure 4.2**). In a preliminary, complete tail-CLIP experiment, we used Z$\alpha\beta$ DAI and we cloned and sequenced some of the plasmids to verify that they contained any insert (data not shown). Having confirmed that we can recover several sequences bound to Z$\alpha\beta$, we decided to create libraries of sequences cross-linked to Z$\alpha\beta$ DAI or its mutant (Mut Z$\alpha\beta$, incompetent of Z-DNA/Z-RNA binding). Unfortunately, we could not observe a clearly distinct band between Z$\alpha\beta$ and Mut Z$\alpha\beta$ DAI in the amplification (**Figure 4.4A**). It was an undesired result but still we decided to perform the pilot next-generation sequencing experiment (MiSeq) to gain insight into the features of nucleic acids captured by Z$\alpha\beta$ or Mut Z$\alpha\beta$. We expected to obtain some specific sequences cross-linked only to the wild-type Z$\alpha\beta$. The second PCR has been performed on the samples shown in **Figure 4.4A** to introduce adaptors complementary to the sequencing chip of the MiSeq instrument. The size distribution of the DNA fragments of the pooled libraries (both WT and Mut Z$\alpha\beta$) is presented in **Figure 4.4B**, for the reference markers of 35 and 10380 bp are included. This distribution has its peak at 189 bp which corresponds to the insert size of 71 bp (length of primers 118, without the extensive poly-A tail).

(A)



(B)

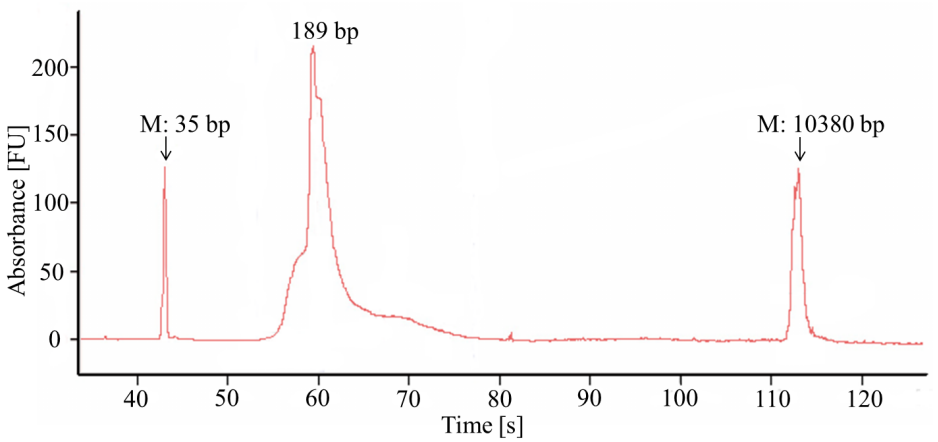


Figure 4.4: The library preparation of RNA cross-linked to Z$\alpha\beta$ DAI or its mutant. The cell lines expressing either Z$\alpha\beta$ or Mut Z$\alpha\beta$ were subjected to the tail-CLIP protocol (see Material and Methods, **Figure 4.2**) **(A)** cDNA (for Z$\alpha\beta$ or its mutant) amplified in the tail-CLIP protocol was separated by the native polyacrylamide gel electrophoresis. **(B)** The size distribution of pooled libraries of sequences recovered from Z$\alpha\beta$ and Mut Z$\alpha\beta$ DAI (assessed by the BioAnalyzer instrument - Agilent). Fragments of 35 and 10380 bp serve as markers (M). The peak of the distribution corresponds to the 189 bp (118 bp primers and 71 bp insert) long DNA fragment.

Profile of the sequences pulled-down with Zα domains

A small fraction of pooled library (1 µl at 2 nM) was used for the sequencing using MiSeq instrument (Illumina). After sequencing, the raw data (reads for forward and reverse primer) have been demultiplexed and a custom script was used to extract insert sequences flanked by the poly-A tail and the L3 adaptor. **Table 4.1** summarizes the number of sequences obtained for each of the samples (Z$\alpha\beta$ and Mut Z$\alpha\beta$ for both sequencing primers). We did not expect a high number of reads as our sample constituted only a small fraction (<1%) of the sequencing lane.

Table 4.1: The sequencing statistics. A total number of reads obtained for Z$\alpha\beta$ and Mut Z$\alpha\beta$ (from forward and reverse primer) is presented. Abbreviations used: w/ - with; w/o - without.

		All reads	Reads w/ insert	Reads w/o insert	Reads w/o L3 adaptor	Reads w/o poly-A
Mut Z$\alpha\beta$	Forward	8516	5624	568	2251	73
	Reverse	7747	5833	853	847	214
Z$\alpha\beta$	Forward	6592	4352	574	1604	62
	Reverse	6304	5190	910	109	95

Next, we plotted the histogram of insert lengths to verify whether the average size corresponds to our predictions (71 bp). We noted that insert sizes are shorter than expected which most likely reflects a suboptimal poly-A tailing and too high RNase I concentration (**Figure 4.5**).

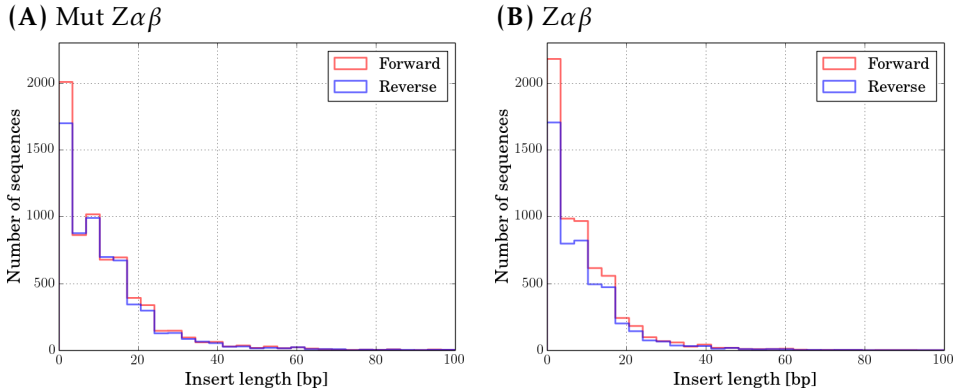


Figure 4.5: Histogram of insert lengths of sequenced libraries. Panel (A) and (B) present histograms of cDNA lengths for Mut Z$\alpha\beta$ and Z$\alpha\beta$ both forward (red) and reverse (blue) primer.

We also examined the basic properties of the inserts (requiring the minimal size of 8 bp), and we compared them between Z$\alpha\beta$ and Mut Z$\alpha\beta$ (**Table 4.2**). We assessed the base composition of the inserts and we noted a slight enrichment for guanosines but both Z$\alpha\beta$ and its mutant had a comparable base content. Z$\alpha\beta$ and Mut Z$\alpha\beta$ differ in their ability to interact with nucleic acids in the left-handed conformation. We were interested whether inserts cross-linked either to WT Z$\alpha\beta$ or its mutant vary in the number of potential Z-forming sequences. For this purpose, we scanned the insert sequences for purine-pyrimidine repeats. We did not detect any differences in the percentage of sequences with Z-forming potential between Z$\alpha\beta$ and Mut Z$\alpha\beta$. It was an indicator that most of the sequences are unspecifically bound to Z$\alpha\beta$ domains or to impurities carried over during sample preparation. Additionally, our protocol (as other iCLIP methodologies) enables to locate the cross-linking site with a high probability. In most cases, the reverse transcription would stop at the position of the peptide which is cross-linked to the RNA molecule (80% of instances) (Huppertz *et al.*, 2014). Thus, the poly-A tailing starts at the position where the reverse transcriptase was stalled. Again, we observed a relatively similar base distribution between Z$\alpha\beta$ and the control at the cross-link site with a slight preference for cytosine. Extraction of the

inserts involved removal of poly-A (seen as poly-T) and L3 adaptor. As a result, we could not distinguish whether thymidine was a part of the tail or the insert so we decided to remove them from analysis and we reported 0 for thymidines at cross-link site.

Table 4.2: Basic sequence composition statistics of the inserts obtained from sequencing of the nucleic acids cross-linked to Z$\alpha\beta$ or Mut Z$\alpha\beta$ DAI.

		Mut Z$\alpha\beta$		Z$\alpha\beta$	
Feature (%)		Forward	Reverse	Forward	Reverse
Base frequency	A	26.3	25.2	24.9	23.7
	T	18.1	19.2	17.5	18.5
	C	24.8	25.3	25.6	26.2
	G	30.8	30.3	32.1	31.7
Potential Z-forming seq (%)		5.1	6.3	5	5.7
Base at cross-linking site	A	34.6	34.3	30.3	30.4
	T	0	0	0	0
	C	34.8	35	37.4	36.3
	G	30.6	30.7	32.3	33.2

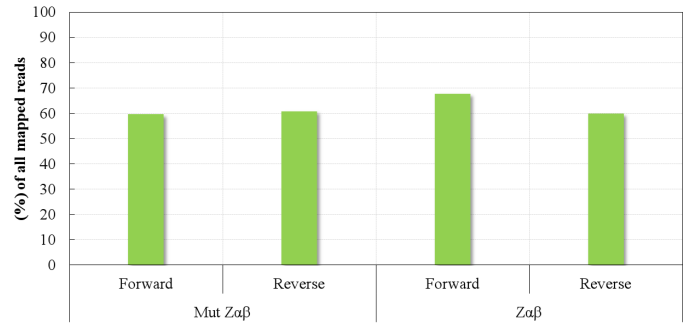
Although the initial analysis of the sequences indicated that nucleic acids pulled-down with Z$\alpha\beta$ or its mutant share many similarities, we filtered and trimmed insert sequences to align them to the human genome (bowtie2). We required at least 20 bp long inserts and only a small fraction (<10%) fulfilled this requirement. The mapping statistics are outlined in **Table 4.3**. The overall mapping rate is higher than 97% with most of the sequences mapping to multiple positions. The explanation why 75% of the reads were equally well mapped to many genomic positions may lie in the fact that most of our sequences are relatively short and/or they were aligned to repetitive regions.

Table 4.3: Statistics of Z$\alpha\beta$ and Mut Z$\alpha\beta$ bound nucleic acids mapped to the human genome.

		Filtered inserts ≥ 20 bp	Uniquely mapped	Multiply mapped	Unaligned	PCR duplicates
Mut Z$\alpha\beta$	Forward	462	151	288	23	19
	Reverse	666	224	422	20	21
Z$\alpha\beta$	Forward	211	56	149	6	4
	Reverse	543	176	350	17	13

In order to assess the percentage of inserts that overlap repeats, we intersected genomic positions of the inserts with annotation of repetitive sequences (the rmsk database). An insert was reported as a repetitive sequence if at least 75% of its length was found to overlap with an annotated repeat. The percentage of inserts overlapping with repeats was around 60% for all samples (**Figure 4.6A**). Additionally, we decided to classify inserts according to the repeat type. **Figure 4.6B** shows that most of the inserts are found within the rRNA repeats. The second enriched fraction is Alu repeats (which belong to SINE - short interspersed elements) and LINE elements. Nevertheless, both Z$\alpha\beta$ and its mutant have similar patterns and with the data obtained we cannot attribute any unique sequences to the wild-type domain. Instead, we found an overrepresentation of ribosomal RNAs (rRNAs, the most abundant RNA species in the cell), which advocates for unspecific interactions.

(A) Overlap with repetitive sequences



(B) Inserts overlapping with repeats classified according to repeat type

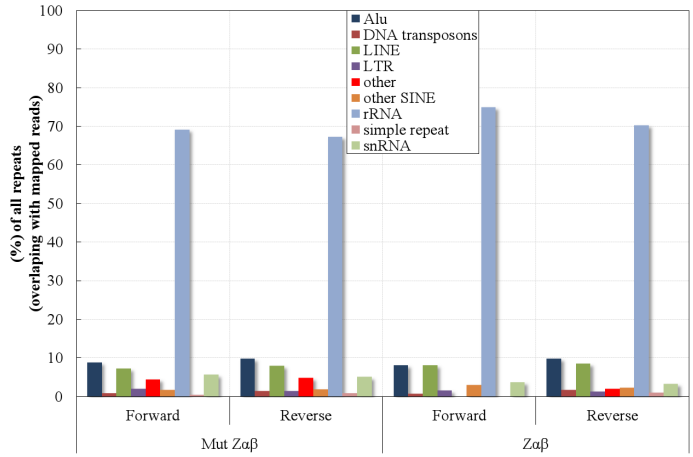


Figure 4.6: Percentage of mapped sequences overlapping with annotated repeats. Minimum 75% overlap between annotated repeat coordinates and genomic positions of mapped reads is required to be included in the analysis. **(A)** The percentage of mapped inserts with at least 75% overlap with annotated repetitive sequences. **(B)** The inserts overlapping with repeats are divided according to the specified repeat type (for each sample i.e. Mut $Z\alpha\beta$ percentages sum up to 100%).

To complete our analysis, we annotated location of each of the mapped inserts to mRNA loci (i.e. coding regions, UTRs, introns) (**Figure 4.7**). Most of the reads fall into intronic regions. The second most represented class is represented by the intergenic regions (no annotation).

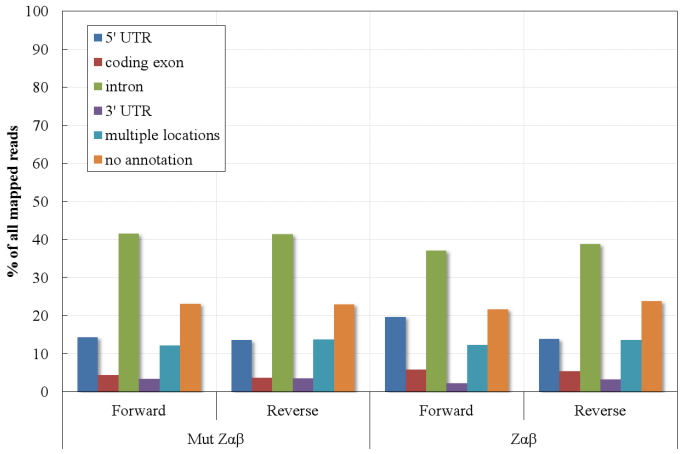


Figure 4.7: Inserts mapping to mRNA loci classified by the functional domain. Each of the inserts mapped to the human genome was intersected with the annotations of each functional part of mRNA (coding region, UTRs, introns etc.). To be counted in a given class a minimum overlap of 75% of length was required.

4.4 Discussion

In this chapter, we confirmed that $Z\alpha\beta$ DAI is enriched in stress granules formed after arsenite treatment (Ng *et al.*, 2013). We provided evidence that $Z\alpha$ domains require Z-DNA/Z-RNA binding activity to reside in SGs. These results were a strong indication that $Z\alpha$ domains can interact with host nucleic acids and are active not only in the context of infection. As stress granules are RNA-protein complexes, we and other reasoned that $Z\alpha\beta$ domains are bound to Z-RNA. We described the development of tail-CLIP protocol which was devised to capture and identify RNA molecules bound to the $Z\alpha\beta$ DAI or its mutant during stress conditions. Unfortunately, we could not detect any specific RNA species associated with $Z\alpha\beta$. We have evaluated whether inserts bound to $Z\alpha\beta$ are enriched in sequences with Z-forming potential compared to sequences associated with Mut $Z\alpha\beta$. Our result indicated that both samples had similar levels of purine-pyrimidine repeats. Moreover, the base content of $Z\alpha\beta$ and mutant inserts were comparable. It was an indication that our preliminary sequencing results represent an unspecific

signal. Nevertheless, to complete the analysis and gain further insight into the nature of RNAs identified in our experiments we mapped inserts longer than 20 bp to the human genome. Most of the sequences with $Z\alpha\beta$ and its mutant were matching equally well multiple genomic locations suggesting a high content of repetitive sequences. Indeed, around 60 % of all sequences overlap with repeats (rmsk annotation). Ribosomal RNA was overrepresented among repeats which could be explained by findings of the previous study indicating the enrichment of ribosomal proteins co-immunoprecipitated with $Z\alpha$ ADAR1 or its mutant (Feng *et al.*, 2011). Alu elements were the second most abundant class of repeats found in our samples - the most abundant retrotransposons in human genome represented by over one million copies (Batzer and Deininger, 2002). In the cytoplasm, Alu elements may be found in 3', 5' UTRs or even in exonic sequences of mRNAs (Fitzpatrick and Huang, 2012). Additionally, polymerase III can independently transcribe Alu elements, which are further processed into small cytoplasmic Alus (scAlu). It is still possible that we recovered Alu elements because both $Z\alpha\beta$ and its mutant are either directly or indirectly linked to co-immunoprecipitated proteins which bind Alu elements. Many additional experiments are required to establish a functional relationship (if any) between Alu elements and $Z\alpha$ domains. One of the important steps would be to perform tail-CLIP with material from a cell line expressing only GFP. It would allow understanding if the signal we observed is related to fully unspecific interactions or has a source in protein complexes co-purified with $Z\alpha\beta$ or its mutant.

Our preliminary sequencing results showed that our approach might be used to recover sequences bound to proteins and obtained material, even with long thymidine runs (from poly-A tailing), can be utilized for the next-generation sequencing. Nevertheless, our tail-clip experiments proved to require further improvements. We need to explore different purification methods, as trials with more stringent washes of GFP beads during immunoprecipitation did not allow distinguishing a $Z\alpha\beta$ from a mutant signal (data not shown). One of the potential strategies could

incorporate an additional purification step with antibodies against a second tag (S-tag in our constructs) or anti-Z$\alpha\beta$ DAI (to be produced). Another consideration is related to the optimization of library preparation. We need to modify the RNase treatment to produce longer RNA fragments because the preliminary library consisted of short inserts. We also need to minimize the length of the poly-A tail by decreasing the time of reaction or/and amounts of adenine added.

Overall, we attempted to characterize nucleic acids interacting with Z$\alpha\beta$ domains. In my opinion, the tail-CLIP methodology, once optimized, may shed light on a nature of RNAs bound by these domains. Moreover, this protocol could be used as an alternative to other Clip strategies providing an on-beads purification strategy preserving almost a single nucleotide resolution.

Acknowledgements

I would like to thank my supervisor, Alekos Athanasiadis, for all his guidance/support and careful reading of this chapter. I am grateful to Luisa Gabriel for her contributions to this project. Finally, I would acknowledge Ewa Chrostek for the reading and comment on this chapter.

Bibliography

Anderson, P. and Kedersha, N. (2008). Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.*, 33(3):141–50. /cited on p. 153/

Athanasiadis, A., Placido, D., Maas, S., Brown, B. A., Lowenhaupt, K., and Rich, A. (2005). The crystal structure of the Zbeta domain of the RNA-editing enzyme ADAR1 reveals distinct conserved surfaces among Z-domains. *J. Mol. Biol.*, 351(3):496–507. /cited on p. 152/

Batzer, M. A. and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.*, 3(5):370–9. /cited on p. 173/

de Rosa, M., Zacarias, S., and Athanasiadis, A. (2013). Structural basis for Z-DNA binding and stabilization by the zebrafish Z-DNA dependent protein kinase PKZ. *Nucleic Acids Res.*, 41(21):9924–33. /cited on p. 152/

Deigendesch, N., Koch-Nolte, F., and Rothenburg, S. (2006). ZBP1 subcellular localization and association with stress granules is controlled by its Z-DNA binding domains. *Nucleic Acids Res.*, 34(18):5007–20. /cited on p. 153, 160/

Feng, S., Li, H., Zhao, J., Pervushin, K., Lowenhaupt, K., Schwartz, T. U., and Dröge, P. (2011). Alternate rRNA secondary structures as regulators of translation. *Nat. Struct. Mol. Biol.*, 18(2):169–76. /cited on p. 153, 173/

Fitzpatrick, T. and Huang, S. (2012). 3'-UTR-located inverted Alu repeats facilitate mRNA translational repression and stress granule accumulation. *Nucleus*, 3(4):359–69. /cited on p. 173/

Flynn, R. A., Martin, L., Spitale, R. C., Do, B. T., Sagan, S. M., Zarnegar, B., Qu, K., Khavari, P. A., Quake, S. R., Sarnow, P., and Chang, H. Y. (2015). Dissecting noncoding and pathogen RNA-protein interactomes. *RNA*, 21(1):135–43. /cited on p. 162/

Ha, S. C., Kim, D., Hwang, H.-Y., Rich, A., Kim, Y.-G., and Kim, K. K. (2008). The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 105(52):20671–6. /cited on p. 152/

Ha, S. C., Lokanath, N. K., Van Quyen, D., Wu, C. A., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K. K. (2004). A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 101(40):14367–72. /cited on p. 152/

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. *Methods*, 65(3):274–87. /cited on p. 162, 168/

Kaiser, W. J., Upton, J. W., and Mocarski, E. S. (2008). Receptor-interacting protein homotypic interaction motif-dependent control of NF-kappa B activation via the DNA-dependent activator of IFN regulatory factors. *J. Immunol.*, 181(9):6427–34. /cited on p. 154/

Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M. J., Scheuner, D., Kaufman, R. J., Golan, D. E., and Anderson, P. (2005). Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.*, 169(6):871–84. /cited on p. 153/

Kuś, K., Rakus, K., Boutier, M., Tsigkri, T., Gabriel, L., Vanderplasschen, A., and Athanasiadis, A. (2015). The Structure of the Cyprinid herpesvirus 3 ORF112-Zα·Z-DNA Complex Reveals a Mechanism of Nucleic Acids Recognition Conserved with E3L, a Poxvirus Inhibitor of Interferon Response. *J. Biol. Chem.*, 290(52):30713–25. /cited on p. 153, 160/

Li, H., Xiao, J., Li, J., Lu, L., Feng, S., and Dröge, P. (2009). Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1. *Nucleic Acids Res.*, 37(8):2737–46. /cited on p. 152/

Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).*, 349(6252):1115–1120. /cited on p. 153/

Mannion, N. M., Greenwood, S. M., Young, R., Cox, S., Brindle, J., Read, D., Nellå ker, C., Vesely, C., Ponting, C. P., McLaughlin, P. J., Jantsch, M. F., Dorin, J., Adams, I. R., Scadden, A. D. J., Ohman, M., Keegan, L. P., and O'Connell, M. A. (2014). The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.*, 9(4):1482–94. /cited on p. 153/

Ng, S. K., Weissbach, R., Ronson, G. E., and Scadden, A. D. J. (2013). Proteins that contain a functional Z-DNA-binding domain localize to

cytoplasmic stress granules. *Nucleic Acids Res.*, 41(21):9786–99. /cited on p. 153, 154, 160, 172/

Pham, T. H., Kwon, K. M., Kim, Y.-E., Kim, K. K., and Ahn, J.-H. (2013). DNA sensing-independent inhibition of herpes simplex virus 1 replication by DAI/ZBP1. *J. Virol.*, 87(6):3076–86. /cited on p. 154/

Rothenburg, S., Schwartz, T., Koch-Nolte, F., and Haag, F. (2002). Complex regulation of the human gene for the Z-DNA binding protein DLM-1. *Nucleic Acids Res.*, 30(4):993–1000. /cited on p. 154/

Schwartz, T. (1999). Crystal Structure of the Z Domain of the Human Editing Enzyme ADAR1 Bound to Left-Handed Z-DNA. *Science (80-.).*, 284(5421):1841–1845. /cited on p. 152/

Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U., and Rich, A. (2001). Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.*, 8(9):761–5. /cited on p. 152/

Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K., Ohba, Y., and Taniguchi, T. (2007). DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature*, 448(7152):501–5. /cited on p. 154/

Upton, J. W., Kaiser, W. J., and Mocarski, E. S. (2012). DAI/ZBP1/DLM-1 complexes with RIP3 to mediate virus-induced programmed necrosis that is targeted by murine cytomegalovirus vIRA. *Cell Host Microbe*, 11(3):290–7. /cited on p. 154/

Wang, Z., Choi, M. K., Ban, T., Yanai, H., Negishi, H., Lu, Y., Tamura, T., Takaoka, A., Nishikura, K., and Taniguchi, T. (2008). Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 105(14):5477–82. /cited on p. 154/

5

Early work aiming to uncover the specificity determinants of A-to-I RNA editing

Abstract

ADAR1 belongs to the enzyme family of adenosine deaminases acting on dsRNA (ADARs). This protein, like other ADARs, is modular and contains a catalytic domain and three (in human) dsRNA binding domains (dsRBD). Two isoforms of ADAR1 have been described: p110 and p150. The longer p150 is expressed under the control of an interferon-induced promoter, whereas p110 is constitutively transcribed. The ADAR1 p150 protein possesses two regions with Z-DNA/Z-RNA binding domain signature: $Z\alpha$ and $Z\beta$. The shorter p110 ADAR1 isoform has only the $Z\beta$ domain. Both ADAR1 isoforms and other ADARs perform catalytic deamination of adenosine to inosine (A-to-I RNA editing) within mRNAs. Inosine is recognized as guanosine by the cellular machinery and thus, A-to-I RNA editing can change amino acids, alter splicing sites and affect miRNA processing. To perform their catalytic function ADARs require RNA that is at least partially double-stranded. The catalytic mechanism of ADARs was not fully elucidated. The structure of ADAR2 catalytic domain provides hints how the reaction of adenosine deamination is performed. This structure revealed that the catalytic pocket is composed of four critical residues (H394, E396, C451 and C516) which coordinate a zinc ion. It is assumed that ADARs uses a base-flipping mechanism to position the adenine in the catalytic center. Moreover, ADARs require inositol hexaphosphate for proper folding and/or catalysis. A-to-I RNA editing can be promiscuous in perfect dsRNA or site-selective in dsRNA with bulges, mismatches and loops. It was postulated that dsRBDs may play an important role in the recognition of specific sequences selected for site-specific editing. Currently, our understanding of specificity determinants of A-to-I RNA editing is relatively limited and it is not possible to predict which bases the enzyme can modify and to which extend. Moreover, the impact of point mutations in secondary RNA structures on the editing efficiency was studied only in few cases. Here, we aimed to study ADAR1 editing mechanism by means of structural biology. Our first goal was to create a heterologous expression system for protein production from several ADAR1 catalytic domain constructs and use

this material for crystallization trials with substrate RNAs. Unfortunately, expression of all of our catalytic domain constructs resulted in insoluble or undetectable proteins. However, we purified and performed initial biochemical characterization of the first and the third dsRBD from human ADAR1 (dsRBD1, dsRBD3). In the conditions tested we did not observe the interaction of dsRBD1 with dsRNA. On the contrary, dsRBD3 was capable of RNA binding but the protein/RNA complex showed decreased solubility. On a different level, we studied RNA editing alterations linked to cancer mutations. In this context, we have performed genome-wide computational prediction of RNA secondary structure of sequences surrounding known cancer silent point mutations. Subsequently, we intersected potential secondary structures with silent point mutations with known A-to-I RNA editing sites. We have compiled a list of known editing sites found in the RNA secondary structures with silent point mutations. Such knowledge can be used to devise *in vitro/in vivo* assays to test the impact of silent point mutations on editing efficiency.

Publication

Unpublished results authored by Krzysztof Kuś and Alekos Athanasiadis.

Contribution

Alekos Athanasiadis and Krzysztof Kuś designed the study and analyzed the results. Krzysztof Kuś performed expression trials, gel filtration experiments and created a computational pipeline to predict secondary structures surrounding silent point mutations.

5.1 Introduction

Adenosine-to-inosine (A-to-I) RNA editing is one of the post-transcriptional modifications of mRNA. ADARs are enzymes catalyzing the deamination of adenosine to inosine in the context of double-stranded RNA. The product of adenosine deamination is inosine, which is recognized as guanosine by the cellular machinery and A-to-I RNA editing can affect many biological processes: recoding of amino acids, removal or introduction of splicing sites, processing of miRNA (Nishikura, 2010). ADARs are modular and consist of a conserved C-terminal catalytic domain and a variable number of dsRNA binding domains (dsRBD). Some structural features are specific to particular ADAR family members: ADAR1 has Z-DNA/Z-RNA binding domains (Zαs) and ADAR3 contains R-domain (arginine-rich single-stranded binding domain). ADAR1 is a multifunctional protein, expressed from two promoters: constitutive (p110 isoform) and interferon-induced (p150 variant). The longer p150 ADAR1 form differs from p110 by one Z-DNA/Z-RNA binding domain (Zα domain). In fact, the prototypic Zα domain was identified as part of ADAR1.

ADARs can modify up to 50% of the adenines in the long, perfect dsRNA (promiscuous editing). Promiscuous editing (ADAR1) emerges as a mechanism inhibiting auto-activation of cytoplasmic dsRNA sensors (MDA5) by host-derived dsRNA (Liddicoat *et al.*, 2015). Mutations in ADAR1 are one of the underlying causes of Aicardi–Goutières syndrome which is an autoimmune inflammatory disorder affecting brain and skin (Rice *et al.*, 2012). A-to-I RNA editing can also be site-selective. Some of the best-studied examples of specific base targeting by A-to-I editing are neurotransmitters i.e. glutamate receptor (GRIA2) (Sommer *et al.*, 1991; Lomeli *et al.*, 1994), serotonin receptor 2C (HTR2C) (Burns *et al.*, 1997) and $GABA_A$ receptor subunit $\alpha 3$ (GABRA3) (Ohlson *et al.*, 2007). It has been postulated that the intramolecular RNA features, like bulges, loops and mismatches are important components of ADAR selective recognition. Additionally, tertiary RNA structure has also been implicated in the ADAR substrate selection process (Ensterö *et al.*, 2009; Tian *et al.*, 2011).

ADARs show only a slight sequence preference manifested by the higher frequencies of some bases neighboring edited adenosine (Lehmann and Bass, 2000).

Currently, little is known about the catalytic mechanism of ADARs. The only available structure of ADAR catalytic domain comes from ADAR2 (Macbeth *et al.*, 2005). This structure revealed critical residues (H394, E396, C451 and C516) involved in the coordination of a zinc ion (found in the catalytic pocket). It is postulated that ADARs perform deamination using a base-flipping mechanism, which would place the adenine in the catalytic center of the enzyme. In addition, a molecule of inositol hexaphosphate (IP_6) was found buried within the protein core and in the proximity of the catalytic center. It was proposed that IP_6 may be important during the protein folding and/or for the deamination reaction (Macbeth *et al.*, 2005).

Another domain shared by all ADARs is the classical dsRBD which is one of the most abundant families of RNA recognition motifs. These domains are found in a wide range of proteins that play a role in the miRNA and mRNA post-transcriptional processing (Masliah *et al.*, 2013). For several years, it was assumed that these domains interact with RNA without any sequence specificity. It has been accepted that dsRBDs recognize only the shape of the double-stranded RNA, as the structural work revealed that the vast majority of the dsRBD-RNA interactions are formed with 2'-hydroxyl groups of the ribose and phosphate backbone. Intriguingly, a study by Stefl *et al.* shed a new light on the recognition of RNA by dsRBDs of ADAR2 (Stefl *et al.*, 2010). These authors investigated the interaction of the first and the second dsRBD of ADAR2 bound to different regions of GluR-2 RNA (encompassing the R/G editing site) (Stefl *et al.*, 2010). Unexpectedly, they discovered that both dsRBD1 and dsRBD2 of ADAR2 form sequence-specific interactions within the minor grove of A-RNA. Moreover, they showed that mutations of residues important for this recognition led to a significant decrease in the ADAR2 editing efficiency. This supported the notion that the base-specific contacts between dsRBD and RNA have a functional importance. Additionally,

the authors tested the impact of mutations of the bases involved in the sequence-specific RNA recognition on the binding affinity of dsRBD. They observed that the changes of the recognition sequence alter the affinity of binding for both dsRBDs (Stefl *et al.*, 2010).

RNA editing is modulated by the features of dsRNA structures (mismatches, bulges and loops). Single point mutations can change the stability of RNA secondary structure (Halvorsen *et al.*, 2010). If a mutation occurs within coding sequence it can either change an amino acid (nonsynonymous substitution) or preserve it (synonymous/silent substitution). The importance of structural features of RNA in editing efficiency was evaluated for few examples only (Dawson *et al.*, 2004; Tian *et al.*, 2011).

The functional role of changes introduced by A-to-I RNA editing was uncovered for some instances. Recoding of glutamate receptors amino acids by RNA editing has a strong impact on their channel permeability (Slotkin and Nishikura, 2013). Changes in the editing levels in the neurotransmitters were associated with several diseases i.e. depression, schizophrenia, amyotrophic lateral sclerosis (Sodhi *et al.*, 2001; Gurevich *et al.*, 2002; Kwak and Kawahara, 2005). Moreover, RNA editing alterations were proposed to play an important role in cancer progression. It was observed that malignant gliomas had a substantially underedited Q/R site of glutamate subunit B (Maas *et al.*, 2001). An increased A-to-I RNA editing of *AZIN1* was found to contribute to the development of human hepatocellular carcinoma. Editing of *AZIN1* introduces a change from serine at position 367 to glycine, and such a change affects the cellular localization of the protein. In addition, an edited AZIN1 variant has an increased affinity to antizyme and inhibits its activity. Antizyme is known to degrade ornithine decarboxylase (ODC) and cyclin D1 (CCND1), which are important players in the regulation of cell cycle. Therefore, antizyme trapped by edited AZIN1 cannot down-regulate an excessive cell proliferation (Chen *et al.*, 2013).

Despite all these advances, our understanding of the substrate

recognition by the catalytic domains of ADARs is still incomplete. As specificity, selectivity, catalytic mechanism and substrate recognition mode are still largely unknown for ADARs, we decided to study them using structural biology and biochemistry. We prepared several constructs for the expression of the catalytic domain of ADAR1 in *Escherichia coli* and/or *Pichia pastoris*. Our aim was to establish a heterologous expression system for the catalytic domain alone or together with dsRBD domains. We planned to use the purified protein for crystallization trials in complex with substrate RNAs. We also planned to use RNAs with analogs of adenine (i. e. 8-azanebularine) that would mimic the intermediates produced during the catalytic cycle in order to study the catalytic mechanism (Maydanovych and Beal, 2006). Therefore, we would block the dissociation of the RNA from the catalytic center and potentially capture the details of the catalytic reaction. Despite the lack of success to produce a functional protein, we present here our strategies and approaches for a future reference.

As mentioned earlier, dsRBDs seem to contribute to the substrate recognition. Most likely both, dsRBDs and catalytic domain contribute to the substrate selection for the deamination. Thus in addition to the catalytic domain, we decided to express dsRBD binding domains of ADAR1. We have performed the initial biochemical characterization of purified domains. Again, we intended to characterize the RNA/dsRBD interactions by means of the X-ray crystallography but due the problems with the complex solubility or lack of detectable interaction with dsRNA we did not reach the desired outcome.

Finally, given the potential importance of RNA editing in cancer evolution, we decided to initiate a study to evaluate the influence of silent point mutations associated with potential RNA secondary structures on editing efficiency. We prepared a computational pipeline which executes large-scale RNA secondary structure prediction of sequences surrounding silent point mutations. We parsed and scored these structures and we intersected the highest scoring ones with the known editing sites (amino acid changing). We produced a list of secondary structures harboring

recoding A-to-I editing site and silent point mutations. These data can be used to check how silent point mutations may control levels of protein variants through A-to-I RNA editing.

In summary, this chapter documents our efforts directed towards understanding the specificity determinants of A-to-I RNA editing with an emphasis on the ADAR1 protein. We describe attempts to obtain proteins through heterologous expression of the constructs with the catalytic domain of ADAR1. We present results of the initial biochemical characterization of the first and third dsRBDs from human ADAR1. Finally, we demonstrate a computational approach to survey predicted RNA secondary structures with silent point mutations for co-occurrence of RNA editing sites.

5.2 Materials and methods

Expression tests of human ADAR1 catalytic domain in *E. coli*

The catalytic domain of human ADAR1 (NCBI: NP_001102 residues 827-1266) was cloned into a pET28a vector with NheI/XhoI restriction enzymes as a His-tag N-terminal fusion protein. The expression tests were carried out in *E. coli* BL21 (DE3). Cells were grown to 0.3 OD and supplemented with an indicated amount of inositol hexaphosphate (IP_6). The growth continued and cell cultures with 0.6 OD were induced with 0.4 mM IPTG. Cells were lysed chemically with BugBuster (Novagen) in the presence of 1 mM PMSF, Benzonase, 2-mercaptoethanol for 1h at 4 °C. Then, soluble fractions were separated from cell debris by centrifugation at 13000 rpm. The soluble and the pellet fractions (resuspended in the loading buffer) were loaded on SDS denaturing polyacrylamide gel and separated. The gel was stained with Coomassie blue (Simply Blue, Invitrogen).

Deaminase containing constructs, *Pichia pastoris* transformation

For extracellular expression in yeast *P. pastoris*, we cloned several constructs into a pPink-α-HC vector (containing α-mating factor from *Saccharomyces cerevisiae*, required for release to the medium) with the StuI restriction enzyme. We prepared two types of inserts encompassing dsRBD and catalytic domain or catalytic domain of ADAR1 only. We decided to generate the expression plasmids for both human and *Strongylocentrotus purpuratus* catalytic ADAR1 domain (spADAR1, the most ancient representative of the ADAR1 family). In our expression trials, we used spD construct (catalytic domain, residues 657-1062 of spADAR1 protein sequence provided in **Appendix**), spD1 (catalytic domain, residues 668-1062 of spADAR1), spRD (dsRBD and catalytic domain, residues 571-1062 of spADAR1). For human ADAR1 (hADAR1), we tested constructs: hD1 (catalytic domain, residues 837-1224 of NP_00110), hRD1 (dsRBD and catalytic domain, residues 719-1224 of NP_00110). As we could not detect any protein expression using pPink-α-HC, we created a plasmid for intracellular expression in *P. pastoris*. The vector backbone was based on a pPink-HC vector and we introduced a Kozak sequence, His-tag and thrombin cleavage site - this plasmid was named pPink-HIS-HC. Using this plasmid we have created and tested the following constructs for the intracellular expression in *P. pastoris*: hRD1, hD (catalytic domain, residues 820-1226 of NP_00110), spD1. Chemically competent *P. pastoris* were prepared according to manufacturer instructions of the PichiaPink Expression System manual using EasyComp kit (Invitrogen). For each transformation around 10 µg of SpeI linearized plasmid was used. Cells were plated on PAD (Pichia Adenine Dropout - without adenine) selection plates. White colonies formed after 4-5 days.

Expression tests in *P. pastoris*

10 ml of BMGY (Buffered Glycerol-complex Medium) in 125 ml flasks was inoculated with a single colony. Cells were grown for 2 days

at 30 °C, 300 rpm. Then, the cells were centrifuged at 1500xg for 5 min and BMGY was decanted. The cells were resuspended in 1 ml of BMMY (Buffered Methanol-complex Medium) to induce expression and were grown for 2 days in the 30 °C incubator, 300 rpm. One of the clones was kept in the BMGY medium as a non-induced control. After the first day of induction, cell cultures were additionally supplemented with 100 µl of 40% methanol. Subsequently, cells were harvested by centrifugation for 10 min at 1500xg. In the case of the extracellular expression tests, the medium was concentrated. The cells were disrupted with YeastBuster (with 1/100 of THP, 1/200 of Benzonase, Complete Mini EDTA-free protease inhibitor cocktail) and mixtures were incubated at room temperature for 15 min. The lysates were centrifuged at 4 °C, 15000xg to separate the soluble and insoluble fractions. Samples from medium, soluble and insoluble cell fractions were used to load onto SDS-PAGE gel (NuPage Bis-Tris 12%, Invitrogen).

Direct PCR screening of *P. pastoris* transformants

Cells from colony or pellets were resuspended in 50 µl of lyticase (100 U/ml in 40 mM K_2HPO_4, pH 7.5). Mixtures were incubated at 37 °C for 30 min and then at 90 °C for 10 min. 5 µl of this solution was used for the PCR with α-For and α-Rev (for pPink-α-HC) or 5'-AOX1 α-Rev (for pPink-HC-HIS vector) primers.

P. pastoris total RNA isolation

Yeast cells (0.5 ml) after 3 h of induction with methanol were harvested by centrifugation at 5000 rpm for 3 min. Non-induced cells were used as a control. Next, 200 µl of lyticase (100 U/ml in 40 mM K_2HPO_4 pH 7.5) was added. The mixture was resuspended and incubated at 30 °C for 40 min, shaking at 300 rpm. Next, solutions were centrifuged at 1000 rpm for 5 min. A total RNA was extracted following manufacturer instructions of RNeasy Mini Kit (Qiagen). The purified RNA was treated with DNAseI (amplification grade, Invitrogen). Samples were incubated at 22 °C for 15 min. Then, 1 µl of 25 mM EDTA was added and reactions

were heated at 65 °C for 10 min and used for PCR.

RT-PCR

cDNA was prepared using SuperScriptIII One-Step RT-PCR System with Platinum Taq (Invitrogen). Reactions were assembled on the ice and contained the primers used for cloning of the inserts. Negative control (RT minus) was included where instead of SuperScriptIII/Platinum Taq mix only Platinum Taq was added. Reverse transcription step was performed at 55 °C for 30 min, followed by the amplification of the PCR product for 40 cycles.

Cloning, expression, purification of dsRBD domains

The first (dsRBD1) and the third (dsRBD3) double-stranded RNA domains of human ADAR1 (NCBI: NP_001102 residues 504-574 and 719-803, respectively) were cloned into a pET28a vector with NheI/XhoI restriction enzymes as a His-tag N-terminal fusion protein. The constructs were expressed in *E. coli* strain BL21 (DE3). Cell cultures with 0.6-0.9 OD were induced with 0.4 mM IPTG. After 3h, cells were harvested by centrifugation (5000 g) at 4 °C. Chemical cell lysis was performed with BugBuster (Novagen) in the presence of 1 mM PMSF, 1 mM $MgCl_2$, cocktail of proteinase inhibitors (Complete Mini, EDTA-free–Roche) and Benzonase (Novagen) for 1h at 4 °C. The protein extract was loaded on a HiTrap IMAC Sepharose FF column (GE Healthcare). The column was washed with 30 mM (dsRBD1) or 125 mM (dsRBD3) imidazole and then the proteins were eluted using a gradient of imidazole 30-500 mM (dsRBD1, start of elution at 50 mM) or 125-500 mM (dsRBD3, start of elution at 200 mM). Only for dsRBD1 the His-tag was cleaved with 10 U Thrombin during an overnight dialysis at 4 °C against MonoS A buffer (50 mM HEPES pH 5.2, 50 mM NaCl). The cleaved protein was further purified by the ion exchange chromatography (MonoS column). The column was washed with ten column volumes of Monos A buffer, the protein was eluted with 50-1000 mM NaCl (pH 5.2) gradient (start of elution at 450 mM NaCl) and the content of the fractions was evaluated

by gel electrophoresis. Buffer exchange and concentration was performed with Amicon-Ultra centrifugal filters (Merck-Millipore) against storage buffer: 50 mM HEPES pH 7.6, 50 mM NaCl, 1 mM $MgCl_2$, 1 mM 2-mercaptoethanol (βME). Proteins were concentrated to 3.1 mM (dsRBD1) and 1.7 mM (dsRBD3).

Gel shift assays

The ability of the purified dsRBDs protein to bind dsRBD was evaluated by gel mobility shift analysis using three different RNA duplex oligonucleotides (CG duplex, Gabra-3 fragment, HTR2C fragment, Integrated DNA Technologies, see **Figure 5.6A**) at different conditions. A mixture of 35 µM protein with 35 µM dsRNA was mixed with glycerol (to the final 12%) and loaded and separated by non-denaturing polyacrylamide gel. The gels were stained with SYBR stain (nucleic acids) followed by Coomassie blue (SimplyBlue, Invitrogen).

Size exclusion chromatography

In order to evaluate the in-solution oligomeric state of the protein, we employed the size exclusion chromatography. An S75 pre-packed column (GE Healthcare) was equilibrated with 50 mM Tris (pH 7.5) and 100 mM KCl (or 1 M KCl) and calibrated using protein standards (molecular mass, 6,500 to 66,000 Da; Sigma-Aldrich). The dsRBD proteins were loaded on the column in a volume of 200 µl and the elution profile was recorded. Based on the calibration curve of standards, we calculated the mass of the each protein and compared to the theoretical value.

Computational survey for the silent point mutations and recoding A-to-I RNA editing in predicted RNA structures

Silent point mutations were selected from the Cosmic database (version 63, 23.01.2013) (Forbes *et al.*, 2008). In order to study potential secondary structure around each of the silent point mutations, we extracted 2500 bp (or the length to the end of the transcript) on both

sides of the mutation coordinate. These sequences were then subjected to local RNA structure prediction software (RNALfold, with -L 1000) (Lorenz *et al.*, 2011). Then, filters were applied to obtain a list of secondary structures where the point mutations were located in the unbranched substructure (stem of at least 8 consecutive base-pairs and a minimum of 12 base pairs; G-U was counted as a base-pair). Next, the coordinates of the structures were intersected with known A-to-I RNA editing site positions deposited in Darned (Kiran *et al.*, 2013) and RADAR (Ramaswami and Li, 2014) RNA editing databases. The final list of structures contained a recoding (the change of the amino acid) A-to-I RNA editing either inside or outside of the substructure with a silent point mutation. The gene tissue expression (based on UniProt annotation) enrichment analysis for the genes found to have coinciding silent point mutations and RNA editing sites were performed on the DAVID web server (Huang *et al.*, 2009).

5.3 Results

5.3.1 Cloning and expression trials of human and *Strongylocentrotus purpuratus* ADAR1 catalytic domains

Human ADAR1 catalytic domain expression trials in *E. coli*

A bacterial heterologous expression system usually offers the highest possible protein yields. Moreover, the growth of *E. coli* is fast and it does not require expensive media. However, the crystal structure of ADAR2 catalytic domain revealed that these domains require inositol hexaphosphate (IP_6) as a cofactor (Macbeth *et al.*, 2005) a molecule absent in bacteria. Although bacteria lack this molecule, we decided to explore whether the IP_6 supplementation before the induction of protein expression could result in the soluble protein production. Therefore, we cloned the catalytic domain of human ADAR1 (encompassing residues 827-1226 of ADAR1 p136, accession: NP_001102, expected molecular

mass: 45 kDa) into a pET28a plasmid and we evaluated protein expression using different IP_6 concentrations. However, supplementation with IP_6 did not result in a proper protein folding and the catalytic domain of ADAR1 was absent in the soluble fraction (**Figure 5.1**) under all IP_6 concentrations tested.

Subsequently, we tested higher IP_6 concentrations (up to 30 mM), but without success (data not shown). One of the potential reasons why the IP_6 complementation did not provide the desired outcome could lie in the fact that IP_6 might have failed to enter the cell. Alternatively, *E. coli* has phytases able to dephosphorylate IP_6 (Greiner *et al.*, 1993). Another reason for the insolubility of human ADAR1 catalytic domain in bacteria could be a lack of specific chaperones or modifications.

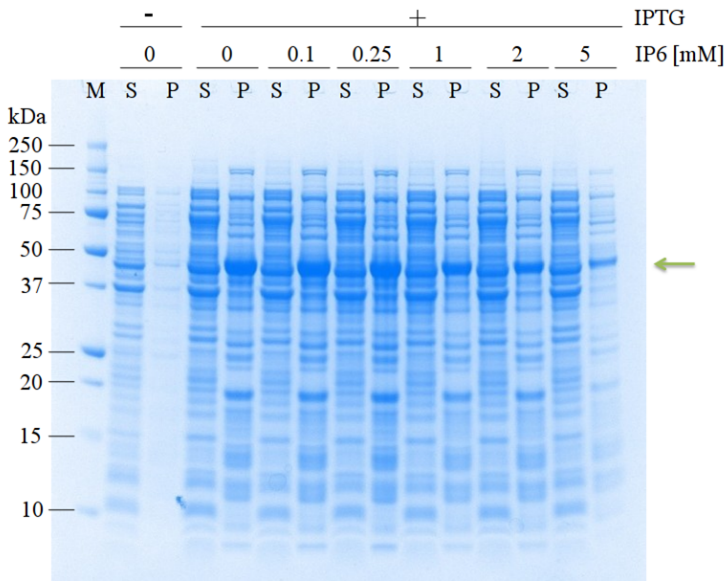


Figure 5.1: The expression test for the catalytic domain human ADAR1 in *E. coli*. NuPage, denaturing, protein gel stained with Coomassie blue. Cell cultures were grown to 0.3 OD and supplemented with indicated concentration of IP_6. When cells reached 0.6 OD, they were induced with 0.4 mM IPTG, a non-induced control is also included. Next, the cells were pelleted and frozen at -20 °C. On the day cell pellets were chemically disrupted, centrifuged to separate soluble and insoluble fractions. The gel was loaded with 10 µl of each sample mixed with loading buffer with 2-mercaptoethanol. Abbreviation used: M - marker, S - soluble fraction, P - pellet (insoluble fraction). The green arrow indicates expected molecular mass of the protein.

Expression tests of catalytic domain ADAR1 constructs in *Pichia pastoris*

We could potentially try to refold the bacterially expressed catalytic domains of ADAR1 but as crystallization requires a homogeneous and good quality protein sample we decided instead to explore the eukaryotic expression systems. We opted for the methylotrophic yeast *P. pastoris*. We selected the PichiaPink Expression System which offers an easy selection of the expression clones. The selection procedure is based on the ADE2 (phosphoribosylaminoimidazole carboxylase) complementation. The strains of PichiaPink are adenine auxotrophs and cannot survive without this base in the medium. The expression plasmids (pPink-α-HC or pPink-HIS-HC) used for cloning contain the ADE2 gene. Transformation of the PichiaPink with pPink plasmid (which integrates into the genome) enables the growth of the cells in the medium without adenine. Additionally, the color of the transformed colonies is an indicator of the relative protein expression. As a first approach, we chose to generate several constructs containing the catalytic domain from ADAR1 in the pPink-α-HC plasmid. This plasmid includes the *Saccharomyces cerevisiae* α-mating factor which directs proteins for export. We prepared constructs with dsRBD and catalytic domain or only catalytic domain both for human (construct names starting with h) and *S. purpuratus* (names starting with sp) ADAR1. At the time we initiated this project, the purple sea urchin was, from the evolutionary point of view, the most ancient species with ADAR1. Therefore, if the project would have proven to be successful we had gained significant insights into the evolutionary history of ADAR1 family. In addition, spADAR1 protein contains only one double-stranded binding domain, which would facilitate the understanding of the contribution of each domain to A-to-I RNA editing. Recently, it has been shown that even sponges (i. e *Amphimedon queenslandica*) possess a protein belonging to ADAR1 family (Grice and Degnan, 2015). **Figure 5.2** illustrates one of the expression trials of human ADAR1 catalytic domain (hD1), sea urchin ADAR1 catalytic domain (spD) and sea urchin dsRBD with catalytic domain (spRD). As negative controls, we used cells transformed with empty plasmid (E) and non-induced cells (of hD1). As we did not observe

expression in the medium, we suspected that the proteins could have been trapped inside cells (either in the soluble or insoluble fraction). Unfortunately, we could not detect any sign of expression in any of the constructs (**Figure 5.2**).

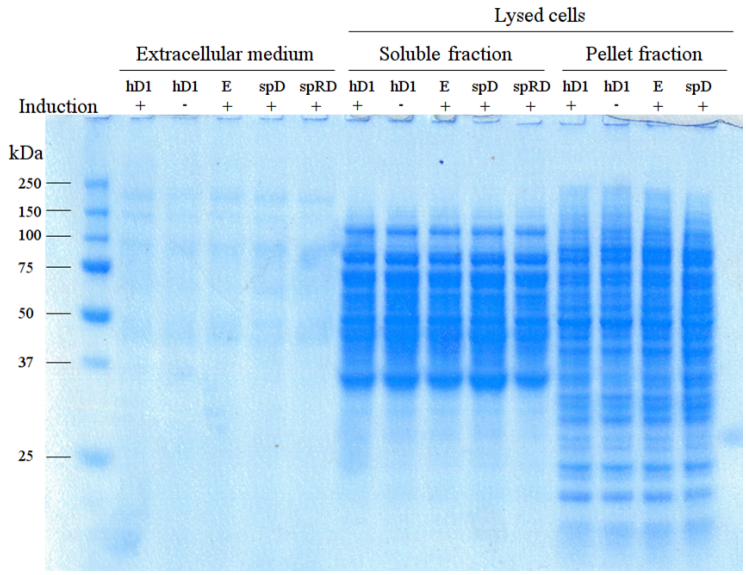


Figure 5.2: Extracellular expression test of different ADAR1 catalytic domain constructs. *P. pastoris* was transformed with different constructs of the human catalytic domain, purple sea urchin catalytic domain and catalytic domain with dsRBD (hD1, spD, spRD respectively, for details, refer to **Materials and methods**). The cells were evaluated for secretion of the proteins of interest (extracellular medium). Additionally, the cells were lysed and both soluble and pellet fractions were analyzed. A cell line transformed with an empty plasmid was a negative control (E). Induction has been performed with methanol.

Although the selection of transformed *P. pastoris* clones is based on growth in medium without adenine and transformants are white instead of pink, we decided to verify if we could amplify the inserts, which would correspond to expected sizes of inserts integrated into the genome of the *P. pastoris*. We could obtain PCR products for tested clones as shown in **Figure 5.3A**. In addition, we isolated total mRNA to verify if we could detect transcripts corresponding to transformed constructs. **Figure 5.3B** demonstrates that hD1 (catalytic domain human ADAR1) is amplified only when we performed reverse transcription of total RNA isolated from methanol-induced cells.

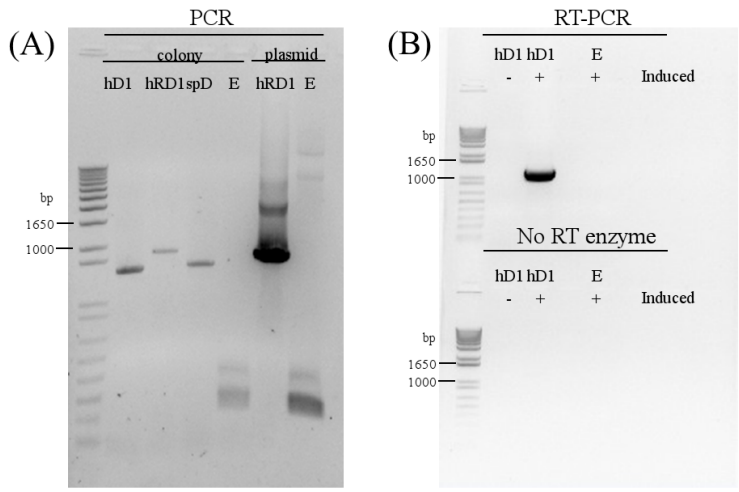
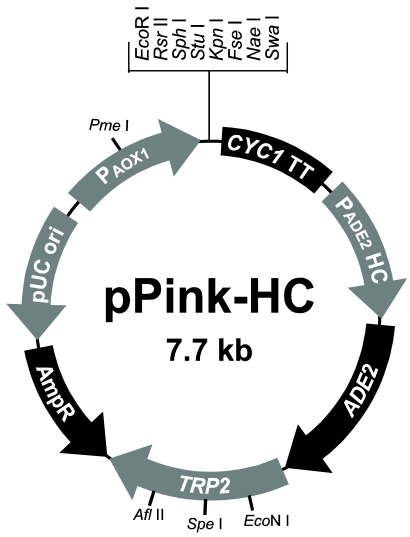


Figure 5.3: Representative results of PCR and RT-PCR analysis of *P. pastoris* transformants. **(A)** Direct PCR on the transformed colonies. *P. pastoris* transformed with: hD1 (catalytic domain human ADAR1), hRD1 (dsRBD3 and catalytic domain human ADAR1), spD (catalytic domain purple sea urchin ADAR1) or E (empty plasmid) were used for screening (panel marked as colony). As positive PCR controls purified empty plasmid (E) or plasmid with dsRBD3 and catalytic domain (hRD1) were used (panel marked as plasmid).
(B) RT-PCR from RNA isolated from induced and non-induced hD1 (catalytic domain human ADAR1) construct. As negative controls experiments were repeated without reverse transcriptase (lower part of the gel, marked as no RT enzyme)

As extracellular expression did not yield any detectable proteins, we decided to prepare constructs which did not contain the secretion signal (**Figure 5.4**). The expression tests without secretion signal would exclude the possibility that protein is trapped and degraded during the export process. The plasmid pPink-HC (**Figure 5.4A**) served as a backbone for the preparation of the vector with His-tag. The insert containing a Kozak sequence (with start codon), His-tag, and thrombin cleavage site was ligated to the pPink-HC plasmid digested with EcoRI and StuI enzymes (**Figure 5.4B**). The ligated fragment contained sequences to recreate the EcoRI and StuI restriction sites.

(A)



(B)

Figure 5.4: Generation of pPink-HIS-HC plasmid for intracellular expression in *P. pastoris*. **(A)** Schematic representation of pPink-HC plasmid. This plasmid lacks the extracellular export signal. Map adapted from PichiaPink Expression System manual. **(B)** Insert introduced to the pPink-HC vector. The fragment consists of two annealed primers and contains Kozak sequence, His-tag and thrombin cleavage site. This sequence was introduced to the EcoRI and StuI double-digested plasmid.

The newly created plasmid was named pPink-HIS-HC and used to clone hRD1 (dsRBD and catalytic domain hADAR1), hD (catalytic domain hADAR1), spD1 (catalytic domain spADAR1). **Figure 5.5** shows the expression trial for human constructs: hD and hRD1. It was found that even when protein is not targeted for secretion we could not detect protein expression. We also performed a Western blot with the antibody against ADAR1, but we did not observe any signal (data not shown). Proteolysis might be one of the underlying reasons for the lack of detectable proteins,

therefore we transformed a strain of *P. pastoris* with knock-out of two proteases (*prb1* and *pep4* genes) but it did not resolve expression problem (data not shown).

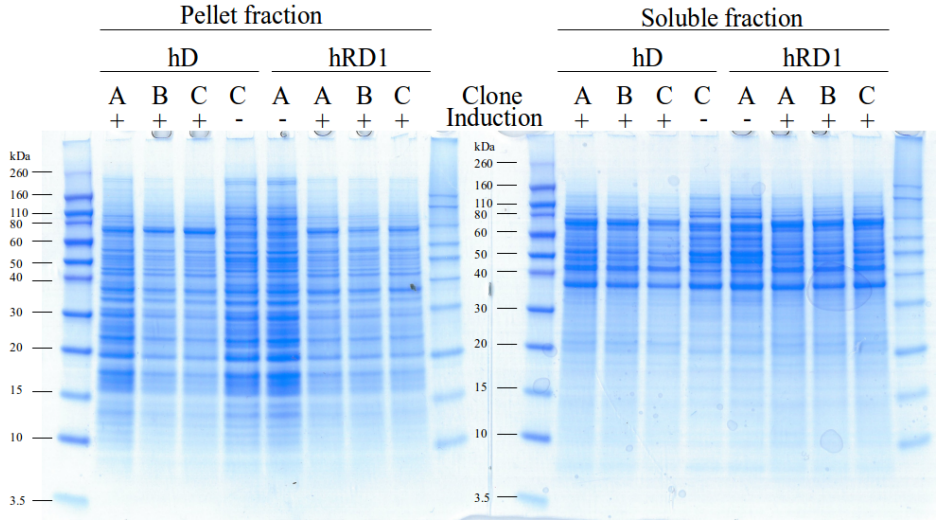


Figure 5.5: Intracellular expression test of human ADAR1 catalytic domain containing constructs. *P. pastoris* cells transformed with hD (catalytic domain of hADAR1), hRD1 (dsRBD and catalytic domain of hADAR1) were grown and induced with methanol (as a control non-induced cells were included). For each construct we evaluated three clones (A, B, C). Next, the cells were lysed and both soluble and pellet fractions were analyzed.

5.3.2 Biochemical characterization of the double-stranded RNA-binding domains of ADAR1

As the expression of the ADAR1 catalytic domain constructs was not successful, we focused our interest on the contribution of the double-stranded RNA binding domains of human ADAR1 in the recognition of editing substrates. First, we tested several constructs with different dsRBD domains of ADAR1 or their combinations (see **Figure 1.11**) either as GST or His-tag fusions. We observed that in all cases substantial amounts of the proteins were found in the insoluble fraction. Insolubility increased with a growing number of dsRBD domains present in the construct (data not shown). As a construct expressing a triple dsRBD were only found in the inclusion bodies, we decided to focus on the optimization of the single

domain expression. We were able to produce and purify to homogeneity the first (dsRBD1) and the third (dsRBD3) dsRBD as His-tagged proteins in *E. coli*. The second dsRBD was poorly soluble and, given that deletion of this domain did not affect the ADAR1 enzymatic activity (at least *in vitro*) (Lai *et al.*, 1995), we did not pursue further trials to improve purification conditions. Our ultimate goal was to characterize protein-RNA interactions by means of X-ray crystallography, so we tested if purified proteins are functional and can bind RNA duplexes. We selected three dsRNA fragments that we used in our assays (**Figure 5.6A**). Two of them constitute fragments of mRNA structures known to be edited both by ADAR1 and ADAR2: Gabra-3 (gamma-aminobutyric acid receptor subunit alpha-3) (Ohlson *et al.*, 2007) and HTR2C (5-hydroxytryptamine receptor 2C) (Burns *et al.*, 1997; Wang, 2000). The Gabra-3 structure is formed within the ninth exon of this gene and represents a relatively short hairpin. As we were concerned with the stability of this duplex we opted for a structure without the loop region. We then included on each end two CG base pairs (**Figure 5.6A**). The structure of HTR2C is formed between sequences in the 3' end of exon 3 and region from intron 3 (Burns *et al.*, 1997). We selected the longest (15 bp), continuous stem from this structure (**Figure 5.6A**). The last dsRNA fragment used in the gel shift assays was a self-complementary, 20 bp CG rich sequence (CG duplex). CG duplex served as a reference and positive control (**Figure 5.6A**). To assess the binding of the RNA duplexes we have utilized the band-shift assays. For binding we tested different salt and pH conditions. **Figure 5.6B** presents the outcome of the dsRBD1 and dsRBD3 migration in the presence or absence of the substrates in a native polyacrylamide gel. Only dsRBD3 interacted with the CG duplex or HTR2C fragment. Nevertheless, we did not observe these complexes entering the gel. Therefore, we checked if precipitation of protein and RNA occurred under the experimental conditions. Indeed, we could detect visible precipitation when we mixed dsRBD3 with either CG duplex or HTR2C fragment.

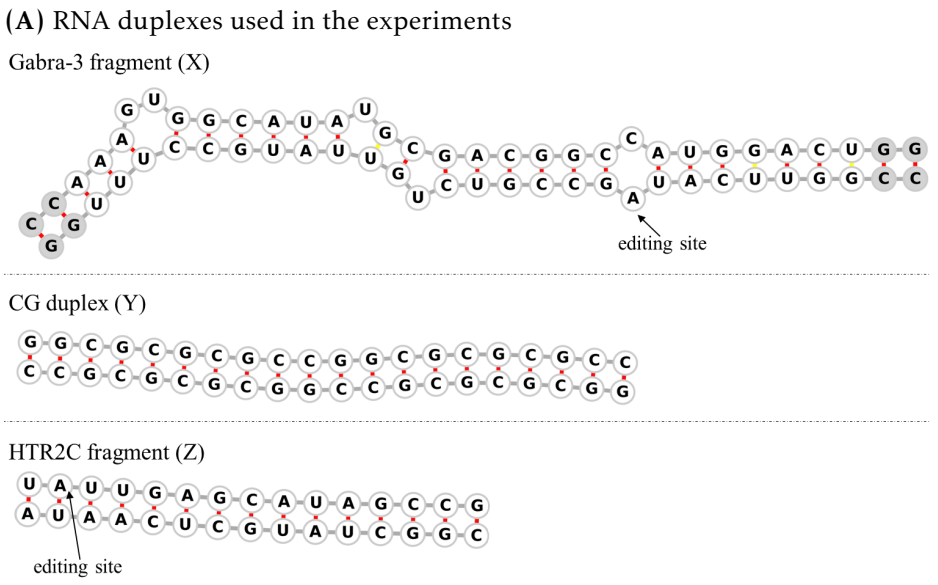


Figure 5.6: Gel shift mobility assays with dsRBDs human ADAR1. (A) Predicted secondary structures of the annealed RNA duplexes used in the gel shift mobility assays. Gabra-3 (abbreviated X) and HTR2C (Z) are fragments of the stems known to be edited by ADAR1. A perfect self-complementary CG rich duplex (Y) was used as a reference. Hydrogen bonding is shown as red or yellow (G-U pair) short lines. Gray shading indicates sequence which is not a part of Gabra-3 stem. **(B)** The band-shift assays for equimolar mixtures of dsRBD1 or dsRBD3 with indicated dsRNAs under different pH conditions. The salt concentration is 50 mM NaCl.

Then, we aimed to understand what the oligomeric states of dsRBDs are. We wanted to check if our sample consisted of aggregates and the addition of the nucleic acids only facilitated the transition to the precipitated form. For this purpose we run several gel filtration experiments. The third dsRBD domain of ADAR1 eluted at 18 kDa, the mass bigger than the theoretically expected 11.8 kDa (**Figure 5.7A, B**). Therefore, we checked by SDS-PAGE whether dsRBD3 in reducing and non-reducing conditions (βME, 2-mercaptoethanol) migrates as single species. As depicted in the **Figure 5.7D**, dsRBD3 does not form efficiently disulfide bonds. Therefore, we also performed gel filtration experiment for dsRBD3 in 1 M KCl (not shown) and the elution time was comparable with low salt conditions supporting the idea that this domain is monomeric in solution. On the other hand, the experiments with dsRBD1 demonstrated that it forms higher molecular assemblies. From the analysis of the protein sample on the SDS-PAGE (in the presence or absence of reducing agent, right side of **Figure 5.7D**), we observed that protein can form disulfide bonds. The gel filtration analysis of dsRBD1 showed one peak of 65 kDa (theoretical 8.4 kDa). The presence of high salt could not destabilize dsRBD1 complex (**Figure 5.7C**).

Dimerization/oligomerization of dsRBD1 may explain why we did not observe dsRBD1 binding to any RNA duplex. We verified that dsRBD1 did not bind HTR2C duplex fragment even under stronger reducing condition (βME) (**Figure 5.8A**, right side). As we observed a precipitation of dsRBD3 with the HTR2C fragment or CG duplex, we explored different buffer compositions trying to avoid formation of aggregates. In one of the experiments we evaluated if addition of salt facilitates formation of soluble complexes that would enter the gel. For this purpose we employed the band-shift assay and assessed the interaction between HTR2C fragment and dsRBD3. Unfortunately, in the highest salt concentration we noticed a release of free RNA oligonucleotides but complexes were still trapped in the well (**Figure 5.8A**, left side). To test the quality of one of the oligonucleotides, we evaluated CG RNA duplex interaction with Zα (ORF112 and ADAR1). As shown

in **Figure 5.8B**, we detected that Zα ORF112 and Zα ADAR1 formed complexes with CG duplex. This experiment served as a control to assure that we can detect formation of complexes in the context of double-stranded RNA.

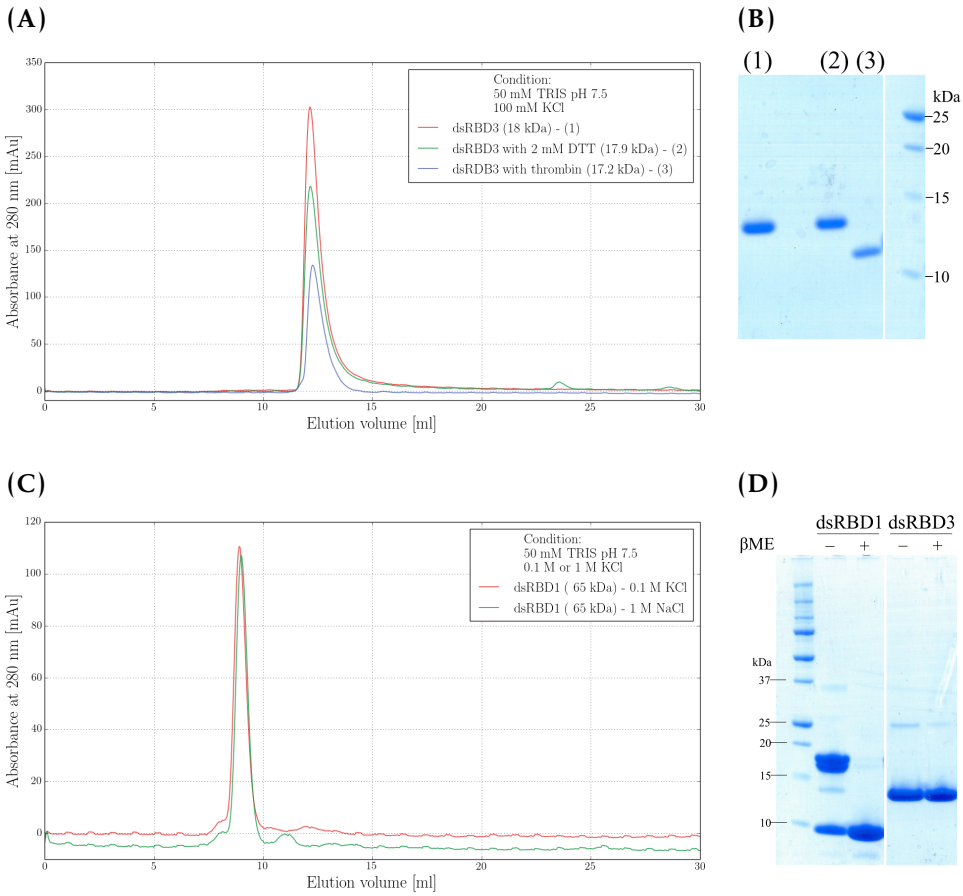


Figure 5.7: Evaluation of the oligomeric state of RBD3 human ADAR1. (A) Gel filtration experiments were performed using RBD3 samples: without DTT (1), with 2 mM DTT (2) or after overnight cleavage with thrombin **(B)** Denaturing, reducing SDS-PAGE separation of proteins collected from (A). Estimated protein size is around 18 kDa which is higher than expected size of monomer dsRBD3 (11.8 kDa). **(C)** The elution profile of dsRBD1 from size exclusion chromatography in low (0.1 M) or high (1M) salt conditions. The protein appeared as a single peak corresponding to 65 kDa (the monomer is estimated to be 8.5 kDa). **(D)** Denaturing SDS-PAGE separation of dsRBD1 and dsRBD3 samples in reducing or non-reducing conditions (βME - 2-mercaptoethanol). The dsRBD1 domain is able to form disulfide bonds.

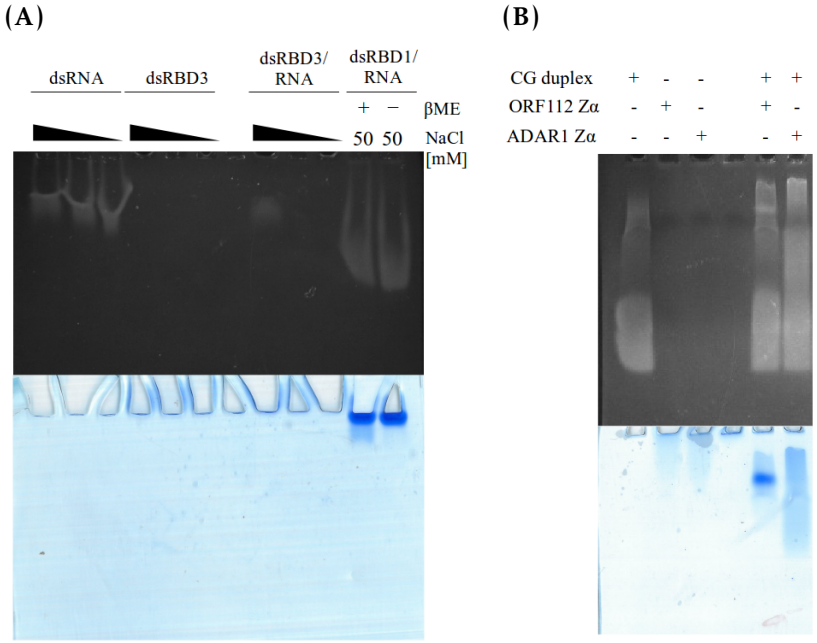


Figure 5.8: Evaluation of the dsRBD1 and dsRBD3 complex formation with HTR2C RNA under different conditions. **(A)** dsRBD3 has been mixed with HTR2C duplex fragment and decreasing salt concentrations were added (500, 250, 125 mM NaCl). Under the conditions tested we did not observed any complex entering the gel. Aggregates were retained in the wells. On the left side, dsRBD1 (shown to form disulfide bonds) was challenged with HTR2C duplex either with or without additional reducing agent (10 mM βME). **(B)** As CG RNA duplex contains sequences with Z-forming potential we utilized this oligonucleotide to verify the interaction with Zα domains (ORF112 or ADAR1). For Zα ORF112 we detected a distinct band of the complex, whereas Zα ADAR1 was found to form diffuse RNA-protein assemblies. The reaction was performed in buffer with 20 mM NaCl and loaded with TBE loading buffer.

In conclusion, we were able to obtain dsRBD proteins suitable for crystallization trials. Nevertheless, we could not observe interaction between dsRBD1 and RNA whereas for dsRBD3 interaction with RNA resulted in precipitation. In addition, we evaluated the oligomeric states of both domains. The construct for dsRBD3 most likely is monomeric in solution. Gel filtration of dsRBD1 indicated that this domain form oligomeric assemblies. DsRBD1 may from covalent dimers (as shown in reducing and non-reducing SDS-PAGE gel).

5.3.3 A computational survey for the consequences of silent cancer mutations on mRNA editing.

As mentioned earlier, A-to-I RNA editing is modulated by structural features of imperfect double-stranded RNA (i.e. loops, bulges). Silent point mutations do not change encoded amino acids but may alter local RNA structure. We wanted to explore whether silent cancer mutations may have functional consequences through alteration of RNA editing of target genes. In this context, we explored on a genome-wide level, whether we can find secondary structures with silent point mutations that have known editing sites in their proximity. We considered this type of information helpful to design *in vitro* assays to provide evidence for synonymous point mutations modulating RNA editing levels. First, we downloaded and analyzed data concerning cancer point mutations which are deposited in Cosmic database (Forbes *et al.*, 2008). In **Figure 5.9A** we plotted a number of reported silent and nonsynonymous mutations found in cancers according to strand and chromosome. This version of database had reported 569050 nonsynonymous and 142039 silent point mutations. The lower number of synonymous than nonsynonymous mutations may reflect a cancer research interest in changes that alter protein sequence. It can be observed that for some chromosomes the number of nonsynonymous mutations exhibited strand bias which may be explained by overrepresentation of the most studied cancer-associated genes (like p53, chromosome 17, minus strand) for which we found 24413 entries. We have also studied the types of substitutions reported in Cosmic database (**Figure 5.9B**). Some of the substitutions are more frequent than others and may be related to the spontaneous deamination of 5-methylcytosine in CpG dinucleotides (Cooper *et al.*, 2010) or may be induced by APOBEC family (Burns *et al.*, 2013; Roberts *et al.*, 2013).

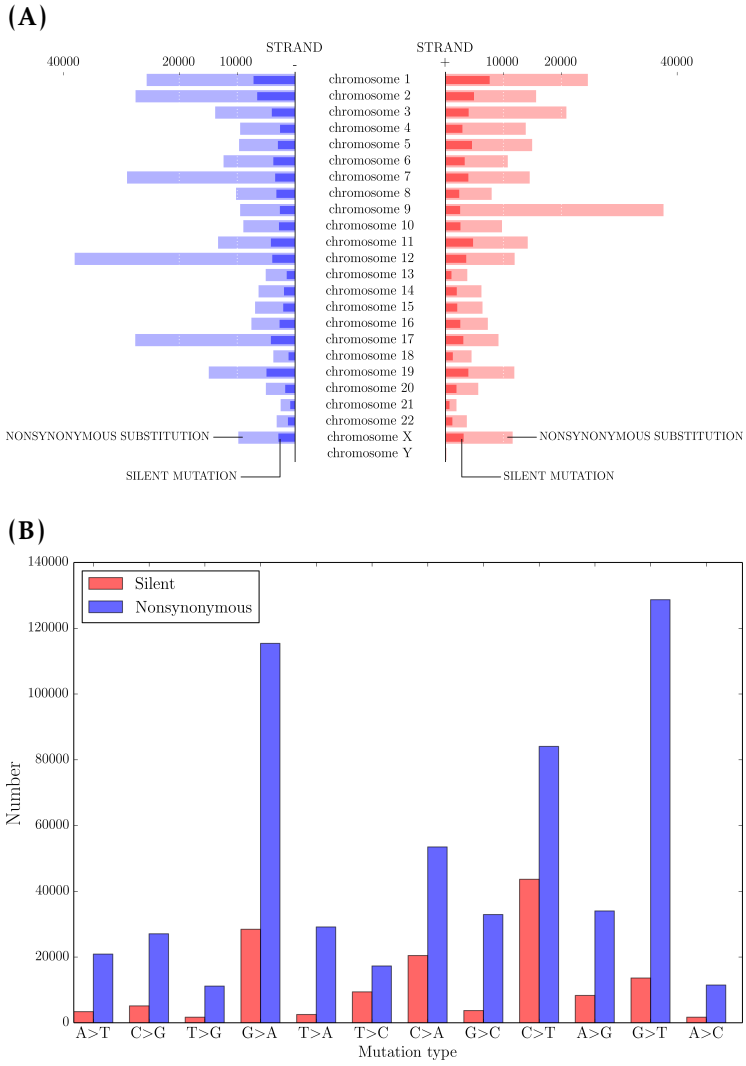


Figure 5.9: Mutations reported in the Cosmic database. (A) A number of silent or nonsynonymous mutations according to chromosome and reported strand in Cosmic database (version 63). **(B)** Types of silent/nonsynonymous substitutions based on Cosmic database.

Next, we extracted 2500 bp (or till end of transcript) around each of the silent point mutations and we performed prediction of the local RNA structure using RNALfold. We parsed and scored each predicted structures. We kept the structures for further analysis if the silent point mutation was located in the unbranched substructure with minimum 12 paired bases (at least 8 bp of continuous dsRNA). This requirement

was based on the fact that some dsRBD can bind substrates with double-stranded region longer than 8 bp (Ramos *et al.*, 2000). We created a two dimensional histogram summarizing dependency the frequency of structures with silent point mutations (**Figure 5.10**). As we expected most of silent point mutations were located in substructures with short stems. On the other hand the distribution of the number of base pairs is wider than the distribution of longest stem sizes in substructures (**Figure 5.10**).

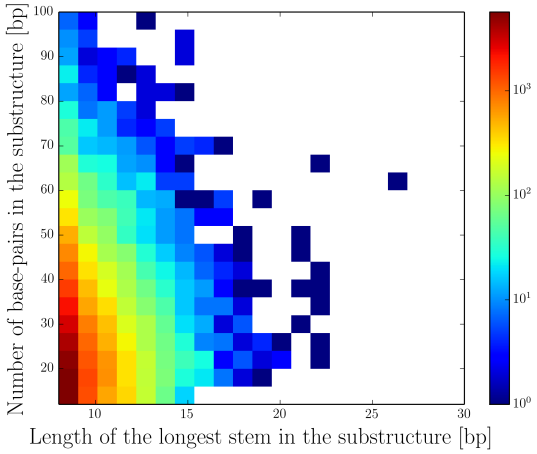


Figure 5.10: Features of potential secondary structures surrounding a silent point mutation. Structure was selected only if a silent point mutation was located in the unbranched substructure with a minimum of 8 bp of continuous double-stranded region (longest stem) and the substructure had at least 12 paired bases. This two-dimensional histogram shows the dependency between the stem size (the longest, continuous base-paired region) and a number of base-pairs in selected substructures.

Ultimately, we intersected the coordinates of the predicted structures with silent point mutations and known recoding A-to-I RNA editing positions. We obtained a list of 553 structures which had an editing site and a silent point mutation in proximity (list of mutations and editing sites can be found in **Table A.4** in **Appendix**). We performed tissue enrichment analysis for the unique genes from our list containing silent point mutation and recoding RNA editing site within predicted structure. We used the online DAVID server (Huang *et al.*, 2009) to perform analysis with UniProt expression annotation. We found that genes expressed in epithelium and brain were significantly enriched in our list **Table 5.1**.

It was shown that A-to-I RNA editing is most prevalent in the neuronal tissues (Li and Church, 2013).

Table 5.1: Tissue expression enrichment analysis. The unique genes containing secondary structures with silent point mutations and recoding A-to-I RNA editing site were analyzed by online server DAVID (Huang *et al.*, 2009) using UniProt expression annotation.

Tissue	Number of genes	% of total	p-value (Benjamini)
Epithelium	79	28.4	$1.4 \cdot 10^{-8}$
Brain	154	55.4	$6.9 \cdot 10^{-5}$
Liver	52	18.7	$4.4 \cdot 10^{-3}$

Most of the editing sites (468) were found outside of the unbranched substructure with silent point mutation. Several structures had multiple editing sites and multiple silent point mutations in the same/similar structure. We have selected one interesting example of SMARCA1 transcript to illustrate four reported recoding RNA editing sites and two silent point mutations found in the same substructure (**Figure 5.11**). We have represented the predicted structure surrounding silent point mutation (**Figure 5.11, top**) and we have zoomed out the substructure with two silent point mutations and four editing sites (**Figure 5.11, bottom**). This substructure has 48 bp and the longest stem spans 8 bp. Two silent point mutations would destroy predicted base pairs and could potentially affect editing levels. Gene *SMARCA1* encodes the protein SNF2L which has been implicated in chromatin remodeling (Barak *et al.*, 2004). This protein was shown to play an important role in brain development (Yip *et al.*, 2012). *In vitro* knockdown studies have demonstrated that depletion of SNF2L results in growth inhibition of cancer cell lines (Ye *et al.*, 2009). Currently, we have no information about A-to-I RNA editing of SNF2L gene. Nevertheless, RNA editing introduces a charge changes in the protein (in one of the sites positively charged Lys is replaced by negatively charged glutamic acid; at another position editing changes neutral Gln to positive Arg).

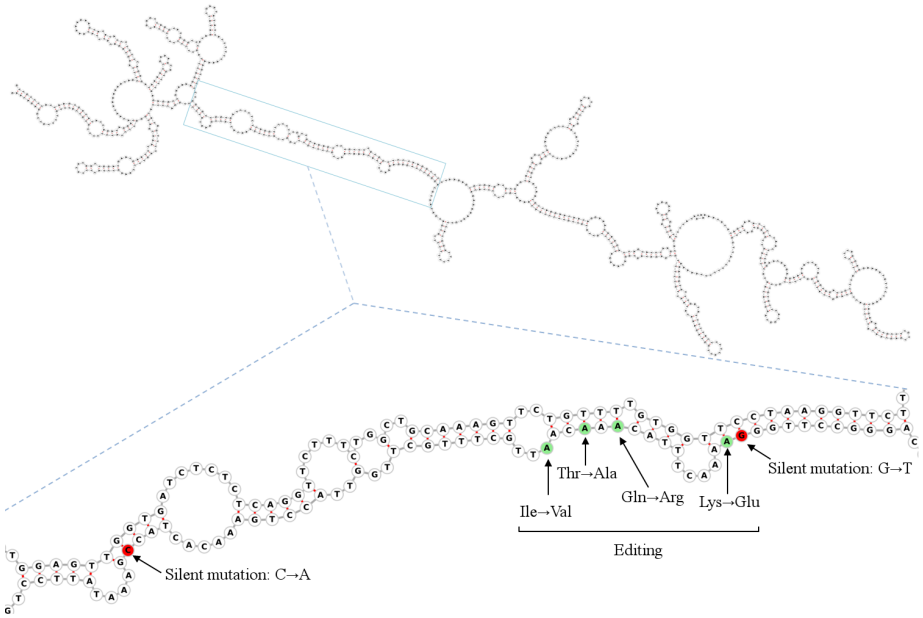


Figure 5.11: Potential RNA secondary structure with silent point mutations and recoding A-to-I RNA editing for SMARCA1 gene. Top of the figure depicts the structure predicted to surround silent point mutations. In the bottom we include zoom in of the substructure containing both silent point mutations (red marking) and editing sites (green color). Indicated substructure contains 48 bp with the longest stem of 8 bp.

In summary, we compiled a list of potential RNA secondary structures surrounding silent point mutations which overlap with recoding A-to-I RNA. Such knowledge may be useful to explore the influence of mutations on the secondary structures which in turn may alter editing patterns.

5.4 Discussion

In this chapter, we described our efforts to create a heterologous expression system for overexpression of ADAR1 catalytic domain containing constructs. We aimed to use these proteins for the structural characterization with substrate and gain insights into the catalytic activity of catalytic domain and mechanism of substrate recognition. We tested a possibility of the protein production in *E. coli* using supplementation with required IP_6 cofactor. Although, the protein was expressed it was present in the insoluble fraction. The failure of the IP_6 supplementation

may be either explained by the fact that this cofactor does not enter the cell or is hydrolyzed by the bacterial phytases. If IP_6 cannot pass through the cell wall an *in vitro* translation system may be a potential solution. Employing the *E. coli* phytase knockouts may also be an option limiting the IP_6 hydrolysis. However, it would require generation of mutated bacterial expression strains. Alternatively, one could consider refolding of the protein from the pellet fraction in the IP_6 presence. On the other hand, we cannot exclude that IP_6 is not the only required molecule. Insolubility of ADAR1 in bacterial cell may be caused by the absence of post-transcriptional modifications or specific chaperones. Proven challenging to obtain a functional protein in *E. coli*, we tried to exploit methylotrophic yeast *P. pastoris*. We tested a series of clones and constructs containing catalytic domain either from human or purple sea urchin ADAR1 without success. We also verified if our constructs were integrated into the genome of *P. pastoris* and we could detect mRNA in the assessed expression clones. It is probable that cells produced trace amounts of the proteins but using our detection system (including the Western blot) we did not ascertain its expression. In this case, one may consider optimization of codon usage to mimic the one utilized in *P. pastoris*. Nevertheless, as a full-length fruit fly ADAR (Ring *et al.*, 2004) and human ADAR2 (O'Connell *et al.*, 1998) were produced in *P. pastoris* based on pPICZ and pSK-FLIS6 system respectively. It is possible that features of human and purple sea urchin ADAR1 (or design of our construct) may be responsible for the lack of protein stability and rapid degradation.

We were interested in understanding how ADAR1 recognizes its substrates as dsRBDs of ADAR2 were shown to engage in sequence-specific binding. We attempted to characterize the interaction between dsRBDs of ADAR1 and RNA. First, we expressed and purified the first (dsRBD1) and the third (dsRBD3) of human ADAR1. We performed several gel filtration experiments to assess the oligomerization state of these domains. dsRBD3 eluted as a peak of around 18 kDa but the experiment in the 1 M KCl did not affect the time taken to migrate

through the column. We concluded that third RBD is monomeric in solution. These results do not agree with the data found in the literature. It is postulated that ADARs (including ADAR1) require dimerization for its activity (Cho *et al.*, 2003). Based on the pull-down experiments with a Flag or His-tagged human ADAR1 p110 it has been shown that dsRBD3 is required for ADAR1 homo-dimerization (Ota *et al.*, 2013). To accommodate our result, dsRBD3 dimerization must occur under specific circumstances. Alternatively, our construct did not cover a region important for dimerization. On the other hand, dsRBD1 most likely engages in higher molecular assemblies which are not affected by 1 M salt. Additionally, dsRBD1 has a tendency to form disulfide bonds. The relevance of these findings awaits further analysis. Nevertheless, we could not show that dsRBD1 is able to interact with the dsRNA. On the contrary, dsRBD3 mixed with dsRNA (HTR2C, CG duplex) resulted in formation a precipitate that prevented us from crystallization trials with RNA. Interestingly, we have not detected the interaction between dsRBD3 and Gabra-3 fragment. Either this duplex (with bulges) is not recognized by dsRBD3 due to insufficient length of the base-paired region or to the lack of the loop region. Overall, our data suggest that the construct of dsRBD3 (residues 719-803) could be suboptimal, as recently a solution structure of the free dsRBD3 was determined by NMR spectroscopy. Curiously, the structure of dsRBD3 revealed an additional, well-structured α helix (α_N) at its N-terminus which is formed by residues 716-726 of ADAR1. This helix was not needed for the interaction with dsRNA (Barraud *et al.*, 2014). It is likely that, as we missed three residues of the α helix, the stability of this region was strongly affected. Binding of RNA by dsRBD3 may induce aggregation leading to precipitation. Therefore, in the future one may consider extending dsRBD3 construct to encompass α_N helix or exclude the entire region.

ADARs (including ADAR1) require dsRNA structures to perform their catalytic activity. We were interested in understanding if silent point mutations may contribute to amino acid changes through A-to-I RNA. We devised a computational pipeline to predict secondary RNA structures

surrounding silent point mutations on the genome-wide level. Such list of filtered structures was intersected with known A-to-I RNA editing sites. We provided a list of interesting examples that can be tested by *in vitro*/*in vivo* assays. *In vitro* studies can be designed to test editing levels in wild-type or mutated (with silent point mutations) structures. The increasing number of transcriptome studies of tumor samples and matched healthy tissue controls could be a valuable source to assess the influence of silent point mutations on the editing levels. Moreover, we presented the predicted secondary structure surrounding two silent point mutations in the SMARCA1 gene. This structure also overlaps with four editing sites that change amino acid and we expect such changes may alter properties of the protein. Nevertheless, one needs to confirm editing of these sites. One could consider producing cell cancer lines with mutated SMARCA1 in the positions of silent point mutations. Such mutants could be prepared with a new genome editing technology CRISPR/Cas9 (Clustered regularly-interspaced short palindromic repeats) (Cho *et al.*, 2013). Editing levels could be compared between mutant and wild-type cell lines and if editing levels were affected by silent point mutations one could try to observe changes of the phenotype.

In summary, we described unsuccessful attempts to characterize on the structural level the catalytic domain of ADAR1. We carried out the preliminary characterization of the dsRBDs of ADAR1. Finally, we produced a list of silent point mutations and known editing sites within the same predicted structures. This knowledge might be used to test whether silent point mutations may introduce variation on the protein level via A-to-I RNA editing.

Acknowledgements

I am deeply grateful to my supervisor, Alekos Athanasiadis, for his guidance, numerous discussions during the work and suggestions for this chapter. I would like to thank Ewa Chrostek for the reading and commenting on the content of this chapter.

Bibliography

Barak, O., Lazzaro, M. A., Cooch, N. S., Picketts, D. J., and Shiekhattar, R. (2004). A tissue-specific, naturally occurring human SNF2L variant inactivates chromatin remodeling. *J. Biol. Chem.*, 279(43):45130–8. /cited on p. 207/

Barraud, P., Banerjee, S., Mohamed, W. I., Jantsch, M. F., and Allain, F. H.-T. (2014). A bimodular nuclear localization signal assembled via an extended double-stranded RNA-binding domain acts as an RNA-sensing signal for transportin 1. *Proc. Natl. Acad. Sci. U. S. A.*, 111(18):E1852–61. /cited on p. 210/

Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E., and Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, 387(6630):303–8. /cited on p. 183, 199/

Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., McDougle, R. M., Brown, W. L., Law, E. K., and Harris, R. S. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, 494(7437):366–70. /cited on p. 204/

Chen, L., Li, Y., Lin, C. H., Chan, T. H. M., Chow, R. K. K., Song, Y., Liu, M., Yuan, Y.-F., Fu, L., Kong, K. L., Qi, L., Zhang, N., Tong, A. H. Y., Kwong, D. L.-W., Man, K., Lo, C. M., Lok, S., Tenen, D. G., and Guan, X.-Y. (2013). Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.*, 19(2):209–16. /cited on p. 185/

Cho, D.-S. C., Yang, W., Lee, J. T., Shiekhattar, R., Murray, J. M., and Nishikura, K. (2003). Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J. Biol. Chem.*, 278(19):17093–102. /cited on p. 210/

Cho, S. W., Kim, S., Kim, J. M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, 31(3):230–2. /cited on p. 211/

Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V., and Chuzhanova, N. A. (2010). Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in

CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum. Genomics*, 4(6):406–10. /cited on p. 204/

Dawson, T. R., Sansam, C. L., and Emeson, R. B. (2004). Structure and sequence determinants required for the RNA editing of ADAR2 substrates. *J. Biol. Chem.*, 279(6):4941–51. /cited on p. 185/

Ensterö, M., Daniel, C., Wahlstedt, H., Major, F., and Ohman, M. (2009). Recognition and coupling of A-to-I edited sites are determined by the tertiary structure of the RNA. *Nucleic Acids Res.*, 37(20):6916–26. /cited on p. 183/

Forbes, S. A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J. W., Futreal, P. A., and Stratton, M. R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.*, Chapter 10:Unit 10.11. /cited on p. 191, 204/

Greiner, R., Konietzny, U., and Jany, K. D. (1993). Purification and characterization of two phytases from Escherichia coli. *Arch. Biochem. Biophys.*, 303(1):107–13. /cited on p. 193/

Grice, L. F. and Degnan, B. M. (2015). The origin of the ADAR gene family and animal RNA editing. *BMC Evol. Biol.*, 15(1):4. /cited on p. 194/

Gurevich, I., Tamir, H., Arango, V., Dwork, A. J., Mann, J. J., and Schmauss, C. (2002). Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron*, 34(3):349–56. /cited on p. 185/

Halvorsen, M., Martin, J. S., Broadaway, S., and Laederach, A. (2010). Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.*, 6(8):e1001074. /cited on p. 185/

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, 4(1):44–57. /cited on p. 192, 206, 207/

Kiran, A. M., O'Mahony, J. J., Sanjeev, K., and Baranov, P. V. (2013). Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res.*, 41(Database issue):D258–61. /cited on p. 192/

Kwak, S. and Kawahara, Y. (2005). Deficient RNA editing of GluR2 and neuronal death in amyotropic lateral sclerosis. *J. Mol. Med.*, 83(2):110–20. /cited on p. 185/

Lai, F., Drakas, R., and Nishikura, K. (1995). Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA

editing of glutamate-gated ion channel transcripts. *J. Biol. Chem.*, 270(29):17098–105. /cited on p. 199/

Lehmann, K. A. and Bass, B. L. (2000). Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry*, 39(42):12875–84. /cited on p. 184/

Li, J. B. and Church, G. M. (2013). Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nat. Neurosci.*, 16(11):1518–22. /cited on p. 207/

Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).*, 349(6252):1115–1120. /cited on p. 183/

Lomeli, H., Mosbacher, J., Melcher, T., Höger, T., Geiger, J. R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., and Seeburg, P. H. (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science (80-.).*, 266(5191):1709–13. /cited on p. 183/

Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, 6:26. /cited on p. 192/

Maas, S., Patt, S., Schrey, M., and Rich, A. (2001). Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. *Proc. Natl. Acad. Sci. U. S. A.*, 98(25):14687–92. /cited on p. 185/

Macbeth, M. R., Schubert, H. L., Vandemark, A. P., Lingam, A. T., Hill, C. P., and Bass, B. L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science (80-.).*, 309(5740):1534–9. /cited on p. 184, 192/

Masliah, G., Barraud, P., and Allain, F. H.-T. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.*, 70(11):1875–95. /cited on p. 184/

Maydanovych, O. and Beal, P. A. (2006). C6-substituted analogues of 8-azanebularine: probes of an RNA-editing enzyme active site. *Org. Lett.*, 8(17):3753–6. /cited on p. 186/

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, 79:321–49. /cited on p. 183/

O'Connell, M. A., Gerber, A., and Keegan, L. P. (1998). Purification of native and recombinant double-stranded RNA-specific adenosine deaminases. *Methods*, 15(1):51–62. /cited on p. 209/

Ohlson, J., Pedersen, J. S., Haussler, D., and Ohman, M. (2007). Editing modifies the GABA(A) receptor subunit alpha3. *RNA*, 13(5):698–703. /cited on p. 183, 199/

Ota, H., Sakurai, M., Gupta, R., Valente, L., Wulff, B.-E., Ariyoshi, K., Iizasa, H., Davuluri, R. V., and Nishikura, K. (2013). ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. *Cell*, 153(3):575–89. /cited on p. 210/

Ramaswami, G. and Li, J. B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, 42(Database issue):D109–13. /cited on p. 192/

Ramos, A., Grünert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000). RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.*, 19(5):997–1009. /cited on p. 206/

Rice, G. I., Kasher, P. R., Forte, G. M. A., Mannion, N. M., Greenwood, S. M., Szynkiewicz, M., Dickerson, J. E., Bhaskar, S. S., Zampini, M., Briggs, T. A., Jenkinson, E. M., Bacino, C. A., Battini, R., Bertini, E., Brogan, P. A., Brueton, L. A., Carpanelli, M., De Laet, C., de Lonlay, P., del Toro, M., Desguerre, I., Fazzi, E., Garcia-Cazorla, A., Heiberg, A., Kawaguchi, M., Kumar, R., Lin, J.-P. S.-M., Lourenco, C. M., Male, A. M., Marques, W., Mignot, C., Olivieri, I., Orcesi, S., Prabhakar, P., Rasmussen, M., Robinson, R. A., Rozenberg, F., Schmidt, J. L., Steindl, K., Tan, T. Y., van der Merwe, W. G., Vanderver, A., Vassallo, G., Wakeling, E. L., Wassmer, E., Whittaker, E., Livingston, J. H., Lebon, P., Suzuki, T., McLaughlin, P. J., Keegan, L. P., O'Connell, M. A., Lovell, S. C., and Crow, Y. J. (2012). Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.*, 44(11):1243–8. /cited on p. 183/

Ring, G. M., O'Connell, M. A., and Keegan, L. P. (2004). Purification and assay of recombinant ADAR proteins expressed in the yeast Pichia pastoris or in Escherichia coli. *Methods Mol. Biol.*, 265:219–38. /cited on p. 209/

Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G., Harris, S., Shah, R. R., Resnick, M. A., Getz, G., and Gordenin, D. A. (2013).

An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.*, 45(9):970–6. /cited on p. 204/

Slotkin, W. and Nishikura, K. (2013). Adenosine-to-inosine RNA editing and human disease. *Genome Med.*, 5(11):105. /cited on p. 185/

Sodhi, M. S., Burnet, P. W., Makoff, A. J., Kerwin, R. W., and Harrison, P. J. (2001). RNA editing of the 5-HT(2C) receptor is reduced in schizophrenia. *Mol. Psychiatry*, 6(4):373–9. /cited on p. 185/

Sommer, B., Köhler, M., Sprengel, R., and Seeburg, P. H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*, 67(1):11–19. /cited on p. 183/

Stefl, R., Oberstrass, F. C., Hood, J. L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R. B., and Allain, F. H.-T. (2010). The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell*, 143(2):225–37. /cited on p. 184, 185/

Tian, N., Yang, Y., Sachsenmaier, N., Muggenhumer, D., Bi, J., Waldsich, C., Jantsch, M. F., and Jin, Y. (2011). A structural determinant required for RNA editing. *Nucleic Acids Res.*, 39(13):5669–81. /cited on p. 183, 185/

Wang, Q. (2000). Requirement of the RNA Editing Deaminase ADAR1 Gene for Embryonic Erythropoiesis. *Science (80-.).*, 290(5497):1765–1768. /cited on p. 199/

Ye, Y., Xiao, Y., Wang, W., Wang, Q., Yearsley, K., Wani, A. A., Yan, Q., Gao, J.-X., Shetuni, B. S., and Barsky, S. H. (2009). Inhibition of expression of the chromatin remodeling gene, SNF2L, selectively leads to DNA damage, growth inhibition, and cancer cell death. *Mol. Cancer Res.*, 7(12):1984–99. /cited on p. 207/

Yip, D. J., Corcoran, C. P., Alvarez-Saavedra, M., DeMaria, A., Rennick, S., Mears, A. J., Rudnicki, M. A., Messier, C., and Picketts, D. J. (2012). Snf2l regulates Foxg1-dependent progenitor cell expansion in the developing brain. *Dev. Cell*, 22(4):871–8. /cited on p. 207/

6

Final remarks and future perspectives

Chapter 2 and **3** are focused on the biochemical and structural characterization of the Zα ORF112, a novel member of Z-DNA/Z-DNA binding domain family. We established that Zα ORF112 can recognize nucleic acids in the left-handed conformation. The structure of the free protein uncovered an unexpected dimer formation between two Zα molecules exchanging C-terminal parts. Although we found that domain-swapping is an artifact induced by crystallization conditions it provides several insights. It suggests that the wing region is flexible and either sulfate ions or DNA stabilize it. Zα ORF112 forms dimers in solution containing sulfate ions but monomer is the predominant form indicating that there is an equilibrium between monomeric and dimeric forms. Our structural analysis of Zα ORF112 bound to $T(CG)_9$ showed that the core of the interaction with Z-DNA mediated by Trp, Asn and Tyr is conserved among all known Zα domains. Despite the overall fold of Zα ORF112 being similar to other Z-DNA/Z-RNA binding domains our structure uncovered novel features not observed among other Zα domains. The structure of Zα ORF112 bound to Z-DNA represents indeed a dimer but is stabilized mostly by van der Waals interactions and does not involve domain-swapping. While dimer interaction surface is extensive it is only established in the context of Z-DNA binding as the protein is monomeric in solution in the absence of sulfate ions. In our opinion, the formation of the on-DNA dimer may provide additional stabilization of the complex and impact the transition rate of left-handed to right-handed nucleic acid conformation. In fact, our mutagenesis experiments of the residue located at the dimer interface (S260) support this notion. Thus, the dimer may be a functional unit bound to the left-handed nucleic acids. In other proteins containing two (or three) Zα domains, such stabilization may be provided by covalently linked domains. Another structural variation concerns the location of the wing region. Protein-protein interactions stabilize the wing regions of the Zα dimer. The wing region of Zα ORF112 points away from Z-DNA and this behavior was not reported for any other Zα·Z-DNA structures. It is noteworthy that the continuous helix may influence the position of the wing region. Zα domains recognize Z-DNA/Z-RNA based on the shape as almost all the

contacts are formed with the phosphate backbone of the nucleic acid. The only known direct base contact is formed between tyrosine residue and guanosine in the syn conformation (a CH-π bond) but this amino acid can establish interaction with any other base in the syn conformation (Ha *et al.*, 2009). Surprisingly, Zα ORF112 revealed a novel direct base recognition between Arg258 and guanosine from the opposite Z-DNA strand. It would be interesting to mutate Arg258 or Ser260 residues of Zα ORF112 and evaluate mutant virus infection efficiency in the fish species. Overall, we could observe novel features of Zα ORF112·Z-DNA interaction because we have used the substrate that extends beyond the minimal binding site. As a result, we could capture second strand interactions and observe on-DNA dimerization.

Our structural and biochemical work concerning ORF112 from cyprinid herpesvirus 3 has established that this viral protein contains functional Zα domain. Cyprinid herpesviruses are members of *Alloherpesviridae* family but other *Herpesvirales* are not known to encode for a protein with Zα domain. Previously, the only described viral protein with Zα domain was E3L from poxviruses. Zα E3L was demonstrated to be critical for the full pathogenicity of vaccinia virus in the mouse model (Kim *et al.*, 2003). The observed phenotype required Z-DNA/Z-RNA binding activity of Zα domain. Exchange of Zα E3L by Zα DAI or Zα ADAR1 preserved the lethality of the virus suggesting that these domains are interchangeable. ORF112 and E3L are expressed as immediate-early genes and their products are transcribed in the initial stages of viral infection (Assarsson *et al.*, 2008; Ilouze *et al.*, 2012) supporting the idea that viral Zαs protect viral nucleic acids from detection. We hypothesize that cyprinid herpesviruses and poxviruses may share a similar host evasion mechanism which requires Zα domain for viral replication. Naturally, experimental confirmation of Zα ORF112 involvement in the evasion of the host antiviral responses is crucial and these experiments are in progress. *In vitro/in vivo* experiments with CyHV3 lacking Zα ORF112 would answer whether this domain is required for the viral replication and pathogenicity. Additionally, the virus with mutated Zα ORF112 would

reveal whether Z-DNA/Z-RNA binding activity is important for the viral evasion mechanisms.

As several other genes in cyprinid herpesviruses and poxviruses are shared (i.e thymidylate monophosphate kinase, thymidine kinase and ribonucleotide reductase) (Ilouze *et al.*, 2006) it has been suggested that these viruses may either share a common ancestor or were subjected to horizontal gene transfer. However, our structural and phylogenetic analyses advocate that Zα ORF112 is more similar to Zα PKZ (encoded in genomes of fish species vulnerable to CyHV3 infection). Therefore, Zα ORF112 might be co-opted from the host fish species by the ancestor of cyprinid herpesviruses. Nevertheless, the similarity between Zα ORF112 and Zα PKZ may be a result of high selective pressure as both domains may bind to common nucleic acids. PKZ, like PKR, can *in vitro* phosphorylate eIF2α after binding to cognate sequences (Yang *et al.*, 2011). From our *in vitro* EMSA-based competition assays, we concluded that Zα ORF112 can compete for nucleic acids with Zα PKZ. We speculate that Zα ORF112 might act as an inhibitor of interferon responses in the relevant fish species by hiding/binding viral nucleic acids (**Figure 6.1**), protecting them from recognition by host innate immune sensors (PKZ).

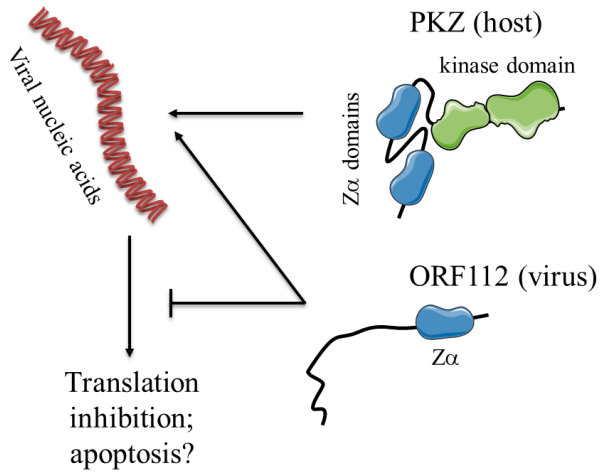


Figure 6.1: Model of inhibition of antiviral responses by ORF112. ORF112 may act as an antagonist of PKZ. Zα ORF112 may bind to viral nucleic acids rendering them inaccessible for PKZ. As a result, PKZ would not be activated and would not be able to evoke translation shut-down and/or apoptosis.

Therefore, we investigated whether the genome of cyprinid herpesvirus 3 contains sequences with potential Z-forming propensity. Indeed, we detected three long GT repeats in the CyHV3 genome and we compared binding of Zα ADAR1 and Zα ORF112 to T$(TG)_8$ repeats. We noted that Zα ORF112 formed only one complex with these repeats whereas Zα ADAR1 was engaged in two complexes. These results could be interpreted as Zα ORF112 being a better stabilizer of GT repeats than Zα ADAR1. It would be interesting to explore whether Zα ADAR1 may substitute Zα ORF112 and what the impact on the viral load and pathogenicity would be. Thus, there are issues which require further studies.

Cyprinid herpesviruses have a dsDNA genome and replicate in the nucleus (Miwa *et al.*, 2007). We have little information about the mechanisms by which DNA of these viruses is exposed to the cytoplasmic milieu and activates cytoplasmic DNA sensors. It is possible that the process of the viral genome deposition in the nucleus is not perfect and due to erroneous release some viral nucleic acids end up in the cytoplasm. In another scenario, the viral capsid proteins may be targeted to proteasomal degradation, liberating viral DNA. Quantification of viral nucleic acids in the cytoplasm could help to distinguish between these scenarios. Further, to fully understand the events in the course of cyprinid herpesvirus infection it would be essential to uncover the nature of viral nucleic acids bound to Zα domains.

Stress granules may function as a bridge between innate immunity and stress responses. SGs formation is connected to translational repression and impacts protein production of invading pathogens. Thus, viruses employ multiple strategies blocking or destabilizing assembly of these structures (as we have observed no stress granules formation for CyHV3). In addition, stress granules may function as a platform to initiate or amplify innate immune responses. Association of multiple molecules involved in interferon responses and cytoplasmic nucleic acid sensors with SGs may contribute to more efficient signal transduction. Yet, it is conceivable that aggregation of nucleic acids and sensors in the

cytoplasmic foci may increase the probability of nucleic acids detection. We have confirmed that Z$\alpha\beta$ DAI is targeted to the stress granules. ORF112 is also found in stress granules (after arsenite treatment). Localization of Zα to stress granules depends on the Z-DNA/Z-RNA binding activity of these domains (as showed for Z$\alpha\beta$ DAI and remains to be demonstrated for Zα ORF112). SGs constitute of mRNAs trapped with ribosomes and mRNA may have double-stranded regions with a propensity to adopt the left-handed conformation. Thus, we predict that Zα domains interact with Z-RNA in stress granules (although we cannot exclude Z-DNA). Z-RNA requires higher energy to be formed than Z-DNA (Brown *et al.*, 2000). One of the biological processes providing force for the transition from the right-handed DNA to the left-handed conformer is a negative supercoiling generated by the moving polymerase during transcription in the nucleus. Yet, the source generating negative supercoiling in the cytoplasm/stress granules remains to be discovered.

One of the questions crucial for a better understanding of the Zα domain function is the origin/source of the nucleic acids bound to Zα domains under the stress conditions. To tackle this problem we developed the tail-CLIP methodology allowing to capture RNA cross-linked to the protein. We used this method to study the two Zα domains of DAI (Z$\alpha\beta$ DAI). We prepared libraries of RNAs associated with Z$\alpha\beta$ DAI or its mutant (Mut Z$\alpha\beta$) and performed preliminary next-generation sequencing of cDNAs. Unfortunately, we could not detect any quantitative differences in the sequences pulled-down with Z$\alpha\beta$ DAI or its mutant. The results indicated that this procedure needs to be improved i.e. by incorporation of additional purification steps and including the control with a cell line expressing GFP only (to test whether our results are pure unspecific background or recovered sequences are bound to proteins associated with Z$\alpha\beta$ domains). We speculate that Alu elements found in our sequencing results may be bound to proteins which co-immunoprecipitated with Z$\alpha\beta$ and its mutant. Moreover, ribosomal RNA was overrepresented in our sequencing data which was not surprising as the previous study provided experimental evidence that Zα domains

associate with ribosomes and have the ability to interact with short rRNA segments (Feng *et al.*, 2011). The authors demonstrated that interaction with rRNA required Z-DNA/Z-RNA binding properties of Zα domains. Nevertheless, considerable amounts of mutant Zα domain were found in the purified ribosomal fraction (Feng *et al.*, 2011). The issues of high rRNA content could be solved if we incorporate additional purification steps with antibodies against Z$\alpha\beta$ or S-tag. Moreover, we may use rRNA depletion to enrich for specific sequences.

ADAR1 is a protein with Zα domains and constitutes an important node of the innate immunity. It emerged as one of the key players in the suppression of interferon responses through A-to-I RNA editing (Mannion *et al.*, 2014; Liddicoat *et al.*, 2015). We tried to understand how ADAR1 recognizes its substrates and catalyzes deamination reaction. We attempted to establish an expression system of catalytic domain constructs in prokaryotic and eukaryotic hosts. Unfortunately, expression trials of catalytic domain of human ADAR1 in *E. coli* yielded protein in the insoluble fraction. During cell growth, we supplemented media with a cofactor (IP_6) required for folding/function of catalytic domains of ADARs. We did not succeed to recover any soluble proteins. In the future, one may consider checking the expression of this domain in the *in vitro* system. Alternatively, if other methods fail, refolding of the catalytic domain of ADAR1 in the presence of IP_6 may be performed. As bacteria lack post-translational modifications, we moved to a yeast expression system (*P. pastoris*). We created a series of human and purple sea urchin catalytic domain ADAR1 constructs. We evaluated extracellular protein production in *P. pastoris*. Moreover, we designed a plasmid without secretion signal to exclude that our proteins are degraded during export to the medium. In both conditions, we did not obtain proteins for further characterization. The lack of detectable protein expression may be a result of poor translation efficiency due to differences in the codon usage. In future experiments, codons of the ADAR1 catalytic domain constructs could be optimized to match *P. pastoris* requirements.

ADAR1 selection of the substrate is a complex process. We aimed to characterize the binding of dsRNA by dsRBDs of human ADAR1. We were able to express and purify dsRBD1 and dsRBD3. Unfortunately, we did not succeed to demonstrate that dsRBD1 interacts with dsRNA. We suspect that dsRBD1 may have sequence preferences and tested oligonucleotides are not cognate substrates for this domain. Alternatively, the formation of high molecular assemblies interferes with binding. We would need to test more constructs to exclude oligomerization as an artifact. On the other hand, we observed that dsRBD3 interacted with dsRNA but the interaction induced precipitation of the complex. One plausible explanation of this phenomenon is a lack of three residues of the α_N helix (due to construct design which may be unstable). We plan to include these three amino acids to cover the full N-terminal helix or remove the entire region.

Another line of research concerning RNA editing by ADARs was meant to investigate how changes in secondary RNA structures may affect editing levels. We focused on the silent point mutations (as they do not change amino acid). We have prepared a computational pipeline that predicts secondary structures surrounding silent point mutations. Next, we intersected dsRNA regions with silent point mutations with known recoding RNA editing sites. We compiled a list of RNA editing sites, RNA secondary structures and silent point mutations occurring in the same regions. This information would allow testing how silent point mutations affect recoding of amino acids. Using this approach we identified SMARCA1 gene which contains four recoding A-to-I editing sites and two silent point mutations in the same secondary structure. We predict that silent point mutations would change editing pattern. One of the important steps will be confirmation of these editing sites in biological samples. Next, we need to perform *in vitro* editing assays of wild-type and mutated (silent point mutation) structures. Finally, it would be interesting to evaluate if changes in editing may have phenotypic manifestations.

Overall, this thesis presents our efforts to understand the biology of the Z-DNA/Z-RNA binding domains and proteins with Zαs (ADARs). We tried to combine several approaches to explore different properties of

Zα domains. We demonstrated that ORF112 from cyprinid herpesvirus 3 contains Zα domain. We solved and described the structure of free and Z-DNA bound form of Zα ORF112. Moreover, we attempted to characterize the identity of nucleic acids bound to Zα domains. We also aimed to address the question of the specificity determinants of the enzymatic reaction (A-to-I RNA editing) carried out by ADAR1 - the first discovered protein with Zα domain. Finally, we generated a list of potential secondary RNA structures surrounding silent point mutations, which also have a recoding RNA editing site in the proximity.

Bibliography

Assarsson, E., Greenbaum, J. A., Sundström, M., Schaffer, L., Hammond, J. A., Pasquetto, V., Oseroff, C., Hendrickson, R. C., Lefkowitz, E. J., Tscharke, D. C., Sidney, J., Grey, H. M., Head, S. R., Peters, B., and Sette, A. (2008). Kinetic analysis of a complete poxvirus transcriptome reveals an immediate-early class of genes. *Proc. Natl. Acad. Sci. U. S. A.*, 105(6):2140–5. /cited on p. 222/

Brown, B. A., Lowenhaupt, K., Wilbert, C. M., Hanlon, E. B., and Rich, A. (2000). The zalpha domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 97(25):13532–6. /cited on p. 225/

Feng, S., Li, H., Zhao, J., Pervushin, K., Lowenhaupt, K., Schwartz, T. U., and Dröge, P. (2011). Alternate rRNA secondary structures as regulators of translation. *Nat. Struct. Mol. Biol.*, 18(2):169–76. /cited on p. 226/

Ha, S. C., Choi, J., Hwang, H.-Y., Rich, A., Kim, Y.-G., and Kim, K. K. (2009). The structures of non-CG-repeat Z-DNAs co-crystallized with the Z-DNA-binding domain, hZ alpha(ADAR1). *Nucleic Acids Res.*, 37(2):629–37. /cited on p. 222/

Ilouze, M., Dishon, A., Kahan, T., and Kotler, M. (2006). Cyprinid herpes virus-3 (CyHV-3) bears genes of genetically distant large DNA viruses. *FEBS Lett.*, 580(18):4473–8. /cited on p. 223/

Ilouze, M., Dishon, A., and Kotler, M. (2012). Coordinated and sequential transcription of the cyprinid herpesvirus-3 annotated genes. *Virus Res.*, 169(1):98–106. /cited on p. 222/

Kim, Y.-G., Muralinath, M., Brandt, T., Pearcy, M., Hauns, K., Lowenhaupt, K., Jacobs, B. L., and Rich, A. (2003). A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, 100(12):6974–9. /cited on p. 222/

Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).*, 349(6252):1115–1120. /cited on p. 226/

Mannion, N. M., Greenwood, S. M., Young, R., Cox, S., Brindle, J., Read, D., Nellå ker, C., Vesely, C., Ponting, C. P., McLaughlin, P. J., Jantsch, M. F., Dorin, J., Adams, I. R., Scadden, A. D. J., Ohman, M., Keegan, L. P., and

O'Connell, M. A. (2014). The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.*, 9(4):1482–94. /cited on p. 226/

Miwa, S., Ito, T., and Sano, M. (2007). Morphogenesis of koi herpesvirus observed by electron microscopy. *J. Fish Dis.*, 30(12):715–722. /cited on p. 224/

Yang, P.-J., Wu, C.-X., Li, W., Fan, L.-H., Lin, G., and Hu, C.-Y. (2011). Cloning and functional analysis of PKZ (PKR-like) from grass carp (Ctenopharyngodon idellus). *Fish Shellfish Immunol.*, 31(6):1173–8. /cited on p. 223/

Appendix

Table A.1: Data collection and refinement statistics for free Zα ORF112.

DATA COLLECTION	
Space group	P $2_12_12_1$
Cell dimensions:	
a, *b*, *c* (Å)	51.7, 54.6, 86.6
α, β, γ (°)	90.0, 90.0, 90.0
Resolution range (Å)	46.18 - 1.75 (1.83 - 1.76)*
R^{**}_{merge}	0.036 (0.17)
Mean I/sigma (I)	13.1 (4.86)
Completeness (%)	94.8 (87.7)
Multiplicity	4.2 (3.7)
Wilson B	31.1

REFINEMENT	
Resolution (Å)	46.18 - 1.76 (1.837 - 1.76)
Number of reflections	21681 (2191)
R^{***}_{work}	0.201 (0.321)
R^{***}_{free} (5%)	0.247 (0.418)
Number of atoms:	
protein	2007
DNA	0
solvent	249
Average B-factors:	
protein	38.40
DNA	-
solvent	45.35
R.M.S. deviations:	
Bond lengths (Å)	0.003
Bond angles (°)	0.677

*Statistics for the highest-resolution shell are shown in parentheses.

** $R_{merge} = \frac{\sum_{hkl}\sum_i |I_i(hkl) - <I(hkl)>|}{\sum_{hkl}\sum_i I_i(hkl)}$

*** $R_{work/free} = \frac{\sum ||F(obs)| - |F(cal)||}{\sum F(obs)}$

Table A.2: Data collection and refinement statistics for Zα ORF112·T(CG)$_9$ complex.

DATA COLLECTION	
Space group	P 3_221
Cell dimensions:	
a, *b*, *c* (Å)	44.82, 44.82, 140.08
α, β, γ (°)	90.0, 90.0, 120.0
Resolution range (Å)	46.69 - 1.5 (1.58 - 1.5)*
R^{**}_{merge}	0.037 (0.68)
Mean I/sigma (I)	22.34 (2.96)
Completeness (%)	99.76 (99.85)
Multiplicity	7.12 (7.34)
Wilson B	22.53

REFINEMENT	
Resolution (Å)	46.69 - 1.5 (1.56 - 1.5)
Number of reflections	26960 (2923)
R^{***}_{work}	0.19 (0.24)
R^{***}_{free} (5%)	0.21 (0.30)
Number of atoms:	
protein	1021
DNA	246
solvent	140
Average B-factors:	
protein	29.40
DNA	24.48
solvent	36.93
R.M.S. deviations:	
Bond lengths (Å)	0.009
Bond angles (°)	1.05

*Statistics for the highest-resolution shell are shown in parentheses.

$$^{**}R_{merge} = \frac{\sum_{hkl}\sum_i |I_i(hkl) - <I(hkl)>|}{\sum_{hkl}\sum_i I_i(hkl)}$$

$$^{***}R_{work/free} = \frac{\sum ||F(obs)| - |F(cal)||}{\sum F(obs)}$$

Table A.3: Thermodynamic parameters for binding of Zα domains to T(CG)$_n$ oligonucleotides. Zα ADAR1 binding to T(CG)$_6$ data was analyzed using a two site model.

Cell (DNA)	Syringe (Zα)	ΔG (kJ/mol)	ΔH (kJ/mol)	-TΔS (kJ/mol)	n	K_d (nM)
T(CG)$_3$	ADAR1	-36.6	-17.8	-18.7	1.4	393
T(CG)$_6$	ADAR1	-38.1	-0.7	-37.5	1.8	210
		-28.9	-21.3	-7.6	1	8700
T(CG)$_3$	ORF112	-35.8	-29.7	-6.1	1.4	538
T(CG)$_6$	ORF112	-37.5	-59.7	22.2	2.4	264
T(CG)$_6$	S260L ORF112	-36.1	-61.9	25.8	2.5	470
T(CG)$_6$	R258A ORF112	-36.9	-48	11.1	2.6	345

Primers used in the tail-CLIP protocol

>L3adaptor
/5rApp/AGATCGGAAGAGCGGTTCAG/3Bio/

>P3_sh
CTGAACCGCTCTTCCGATCT

>MiSeqF_1
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGAGATTTTT
TTTTTTTTTTTT

>MiSeqR_1
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCAAGTCTG
AACCGCTCTTCCGATCT

>MiSeqF_2
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTCAATTTTT
TTTTTTTTTTTT

>MiSeqR_2
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGCTCGCTG
AACCGCTCTTCCGATCT

Strongylocentrotus purpuratus ADAR1 protein sequence (spADAR1)

MGSNTELANGILNKLKESSGSMETLELARSLGRRSRRDVNPTLYRMQ
KNGLTLEVSASPPKWGLKNEVTVGPLHGATAEDGDNLQEAEETAEE
ALSGGSGGMERDDFMDEGSGGDVPVAEPFGHPVNHSASIDAGQLYP
PPLSSYNGTHSVPGVDPMPSPTNISSSSDDSETSEQPVIFAKPPPPPHEL
QQSFAKPYTPNGMDHRLLVALSEKVESIHSNDLAKQLGYSTKKEINPT
LFSMQKKGLVRKVCESPPMWVITPYGRQILETDQQQQSPEQSQQKP
GQFPVMMSSPMPGTAPHIVNFGPHIEQHQGVPGQMFHHVPQSPQPL
LGSNDDRQSIYNERLNLQAAIKAGIPMEFATLMFASPDMEIKVLCALS
NQQTQGTVEIVHNVGLNRGKAKDVNPSLYGLAKRGNITKVTDSPPT
WHINEKGVELINKEKKPIEMNGATKPMEIPNHVNMVTGAPGVLMPQ
PMSSPNHQGGILPPDPRTVLQNWDQGRPGPGFIPVSSFTPTIPTQVAP
IRATTMPYMPPNSFANPASVNSAPIPASPAPQPQSSHSAPVIGSSTGQT
ISNEMFAAINKNPVSALTEYAQARHLPVSIDLLQQSGPPHNPRFVFAA
CVGGRRFQHVTSRSKKDGRREAADMALRTLVAEGSLQPKLTPMPIV
HTSSNGETMFYDRIAALSHQTFNSLAASIPDSISGRKVLAALIMKQGE
DDEGMVISLGTGNRCVTGDKLSMEGRTVNDSHAEIITRRAFLRYLYN
QLQAYAKTPNETILTQGTNGKLRLLPDVSLHLYISTAPCGDGAQFSRT
DAGENEEGPNGIDFCGFAKHLPTFGKTSQGLLRTKMEQGEGTIPVTT
RESVRTWDGIMRGERLRTMSCSDKVASWNLLGLQGALLSHFIEPMYL
SSISLGSLYHHGHLARAVCCRVSSAHDNFTPTELDNLSLPAEYHVNHP
QLGCVRAIDPPRGTEKTKSLSINWYRGCEKPEVTDGTKGRIQGVLHS
QLCKAEMFSEYQQTCRLFDRTDLLESRTYHDAKLSAESYYTAKQYLK
CILHQANYGSWMEKPIEEELFANYPYDVPDYA

Table A.4: Structures with silent point mutations and recoding amino acids A-to-I RNA editing. Column "Where editing" refers to location of editing site: inside or outside substructure with silent point mutation. Abbreviations used: mut - mutation, N° - number, aa - amino acid, pos - position in genome, chr - chromosome, gen - genomic, edit - edited.

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
ZSCAN2	uc002bkr.3	chr15	+	85164478	G	A	10	25	85164858	T	A	outside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116595	K	R	outside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116613	K	R	outside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116632	T	A	outside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116641	R	G	outside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116868	Y	C	outside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116847	K	R	outside
ZNF721	uc003gag.3	chr4	-	436275	G	A	9	36	435906	K	E	outside
ZNF721	uc003gag.3	chr4	-	436275	G	A	9	36	435924	K	E	outside
ZNF721	uc003gag.3	chr4	-	436275	G	A	9	36	435936	I	V	outside
ZNF721	uc003gag.3	chr4	-	435804	G	A	8	43	435936	I	V	outside
ZNF721	uc003gag.3	chr4	-	435804	G	A	8	43	435924	K	E	outside
ZNF646	uc002eap.3	chr16	+	31087947	C	T	8	15	31088361	Y	C	outside
ZNF600	uc002qab.4	chr19	-	53269322	C	T	10	19	53269118	I	V	outside
ZNF600	uc002qab.4	chr19	-	53268920	T	C	9	16	53269118	I	V	outside
ZNF587	uc002qqb.2	chr19	+	58369971	G	A	8	27	58370160	K	R	outside
ZNF584	uc002qsp.3	chr19	+	58928265	G	A	9	51	58928568	K	R	outside
ZNF584	uc002qsp.3	chr19	+	58928265	G	A	9	51	58928567	K	E	outside
ZNF584	uc002qsp.3	chr19	+	58928265	G	A	9	51	58928555	S	G	outside
ZNF584	uc002qsp.3	chr19	+	58928265	G	A	9	51	58928547	K	R	outside
ZNF582	uc002qmy.3	chr19	-	56895309	C	T	11	29	56895723	I	V	outside
ZNF582	uc002qmy.3	chr19	-	56895795	G	A	8	21	56896203	N	D	outside
ZNF582	uc002qmy.3	chr19	-	56895309	C	T	11	29	56895713	Q	R	outside
ZNF582	uc002qmy.3	chr19	-	56895309	C	T	11	29	56895692	K	R	outside
ZNF582	uc002qmy.3	chr19	-	56895309	C	T	11	29	56895686	Y	C	outside
ZNF582	uc002qmy.3	chr19	-	56895840	A	G	9	34	56896203	N	D	outside
ZNF582	uc002qmy.3	chr19	-	56895603	C	T	8	13	56895363	R	G	outside
ZNF582	uc002qmy.3	chr19	-	56895603	C	T	8	13	56895428	Y	C	outside
ZNF582	uc002qmy.3	chr19	-	56895309	C	T	11	29	56895428	Y	C	outside
ZNF582	uc002qmy.3	chr19	-	56895603	C	T	8	13	56895686	Y	C	outside
ZNF582	uc002qmy.3	chr19	-	56895348	G	C	11	28	56895428	Y	C	outside
ZNF582	uc002qmy.3	chr19	-	56895309	C	T	11	29	56895363	R	G	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
ZNF582	uco02qmy.3	chr19	-	56895348	G	C	11	28	56895363	R	G	outside
ZNF43	uco10ecv.3	chr19	-	21992037	T	C	8	13	21991662	T	A	outside
ZNF43	uco10ecv.3	chr19	-	21992037	T	C	8	13	21991679	K	R	outside
ZNF43	uco10ecv.3	chr19	-	21991701	T	C	8	16	21991662	T	A	outside
ZNF43	uco10ecv.3	chr19	-	21991701	T	C	8	16	21991679	K	R	outside
ZNF417	uco02qqq.3	chr19	-	58420208	G	T	8	18	58420430	R	G	outside
ZNF417	uco02qqq.3	chr19	-	58420208	G	T	8	18	58420303	K	R	outside
ZNF323	uco10jra.3	chr6	-	28295208	G	C	9	27	28295174	Q	R	outside
ZNF283	uco020xr.4	chr19	+	44351562	G	A	11	23	44351516	R	G	outside
ZNF283	uco020xr.4	chr19	+	44351562	G	A	11	23	44351535	K	R	outside
ZNF283	uco020xr.4	chr19	+	44351562	G	A	11	23	44351544	E	G	outside
ZNF283	uco020xr.4	chr19	+	44351562	G	A	11	23	44351549	K	E	outside
ZNF267	uco02ecs.4	chr16	+	31926643	C	G	10	15	31926022	Q	R	outside
ZNF233	uco21uvi.1	chr19	+	44777940	G	A	9	29	44778513	D	G	outside
ZNF195	uco10qxr.2	chr11	-	3380815	C	T	8	13	3380388	Q	R	outside
ZNF195	uco10qxr.2	chr11	-	3380758	C	A	8	14	3380388	Q	R	outside
ZIC1	uco03ewe.3	chr3	+	147128621	G	A	9	12	147128798	K	R	outside
ZCCHC9	uco03khk.4	chr5	+	80608438	G	T	12	34	80608441	Q	R	outside
ZBTB48	uco09vmc.2	chr1	+	6648819	G	T	8	21	6648361	T	A	outside
WDR85	uco04cnk.1	chr9	-	140458997	G	A	10	27	140458963	H	R	outside
WDR33	uco02tpg.2	chr2	-	128481959	G	T	8	20	128482009	K	R	outside
WDR20	uco01yld.3	chr14	+	102675625	C	A	12	21	102675409	H	R	outside
WDR18	uco02lqm.1	chr19	+	994261	G	A	8	19	994079	M	V	outside
VPS37B	uco01udl.3	chr12	-	123351890	A	C	9	18	123351949	Y	C	outside
VNN2	uco03qdv.3	chr6	-	133070833	G	A	11	62	133070983	K	E	outside
UTP14C	uco01vgb.3	chr13	+	52604760	C	T	10	17	52604985	Q	R	outside
UTP14C	uco01vgb.3	chr13	+	52604760	C	T	10	17	52604880	Q	R	outside
URGCP	uco03tiz.3	chr7	-	43917153	G	T	9	20	43917518	Q	R	outside
URGCP	uco03tiz.3	chr7	-	43917630	G	A	10	18	43917518	Q	R	outside
UBE4A	uco01psw.3	chr11	+	118235858	C	T	14	24	118235809	E	G	outside
UBAP2	uco03ztq.1	chr9	-	33989028	G	T	9	32	33989112	T	A	outside
TUBA4A	uco10zkz.1	chr2	-	220115496	G	A	9	31	220115135	E	G	outside
TUBA4A	uco10zkz.1	chr2	-	220115496	G	A	9	31	220115403	S	G	outside
TTC17	uco01mxi.3	chr11	+	43464878	G	T	8	18	43464916	N	D	outside
TRIM56	uco03uxr.3	chr7	+	100732026	G	C	8	19	100731402	K	R	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
TRIM56	uco03uxr.3	chr7	+	100731312	G	A	11	31	100731402	K	R	outside
TRIM28	uco02qtg.1	chr19	+	59061389	C	T	9	18	59061154	E	G	outside
TMX2	uco01nlc.2	chr11	+	57507578	G	C	9	20	57507701	N	S	outside
TMEM8A	uco02cgv.4	chr16	-	426264	G	T	10	33	426392	N	S	outside
TMEM131	uco02syh.4	chr2	-	98412841	C	T	8	13	98412706	N	D	outside
TLE4	uco04alc.3	chr9	+	82336718	C	A	8	19	82336658	Q	R	outside
TLE3	uco10bil.1	chr15	-	70351091	C	T	8	71	70351034	T	A	outside
TBC1D8	uco10fiv.3	chr2	-	101654985	G	A	10	12	101655099	M	V	outside
TBC1D8	uco10fiv.3	chr2	-	101655069	C	A	9	13	101655099	M	V	outside
TBC1D1	uco03gtb.3	chr4	+	38020051	C	A	9	17	38020033	K	R	outside
TAX1BP1	uco11jzo.2	chr7	+	27868360	C	T	9	47	27868417	H	R	outside
TAF1L	uco03zrg.1	chr9	-	32635568	C	T	9	65	32635216	Q	R	outside
TAF1L	uco03zrg.1	chr9	-	32635565	C	T	9	65	32635216	Q	R	outside
TACC1	uco03xlz.3	chr8	+	38677328	C	T	8	16	38677553	Y	C	outside
SYT14	uco01hhs.4	chr1	+	210334371	G	A	8	40	210334259	K	E	outside
SYT14	uco01hhs.4	chr1	+	210334353	A	G	8	40	210334259	K	E	outside
SPSB3	uco02cmu.3	chr16	-	1827853	C	T	9	14	1827756	K	R	outside
SPEN	uco01axk.1	chr1	+	16254678	G	A	8	15	16255439	K	E	outside
SPEN	uco01axk.1	chr1	+	16258068	C	T	8	24	16258632	N	S	outside
SPEN	uco01axk.1	chr1	+	16255365	C	T	9	27	16255439	K	E	outside
SOX13	uco01ham.3	chr1	+	204082198	C	T	10	34	204082228	Q	R	outside
SON	uco02yse.1	chr21	+	34923231	G	T	9	47	34922801	T	A	outside
SON	uco02yse.1	chr21	+	34922736	T	A	8	25	34922801	T	A	outside
SLX4	uco02cvp.2	chr16	-	3632609	C	T	11	18	3632632	E	G	outside
SLCO4C1	uco03knm.3	chr5	-	101592850	G	A	9	31	101592931	T	A	outside
SLC26A2	uco03lrh.3	chr5	+	149360445	C	T	8	30	149361011	R	G	outside
SLC16A5	uco02jmr.3	chr17	+	73096144	G	T	10	19	73096609	N	S	outside
SLC16A5	uco02jmr.3	chr17	+	73096372	G	T	8	14	73096609	N	S	outside
SH3BP2	uco03gfi.4	chr4	+	2826420	G	A	11	17	2826853	S	G	outside
SH3BP2	uco03gfi.4	chr4	+	2831394	C	A	8	19	2831318	K	E	outside
SH3BP2	uco03gfi.4	chr4	+	2831382	T	G	8	19	2831318	K	E	outside
SETDB1	uco01evv.2	chr1	+	150900288	G	T	8	17	150900449	R	G	outside
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603125	K	E	outside
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603124	K	R	outside
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603109	N	S	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603107	S	G	outside
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603073	K	R	outside
SEMA4D	uco04aqp.1	chr9	-	91994305	G	A	8	14	91994064	N	S	outside
SDAD1	uco03hje.4	chr4	-	76891501	G	C	13	18	76891486	I	V	outside
SCGB2A1	uco01nta.2	chr11	+	61976227	C	T	8	17	61976208	K	R	outside
SAMD9L	uco03umh.1	chr7	-	92761753	C	T	8	40	92762389	T	A	outside
SAMD9L	uco03umh.1	chr7	-	92762680	T	C	8	23	92762389	T	A	outside
SAMD9L	uco03umh.1	chr7	-	92762653	A	G	8	25	92762389	T	A	outside
SAMD9L	uco03umh.1	chr7	-	92762500	C	T	8	23	92762389	T	A	outside
RYBP	uco03dpe.3	chr3	-	72495720	C	T	8	22	72495657	T	A	outside
RUFY1	uco03mka.1	chr5	+	179028915	G	A	10	31	179028978	H	R	outside
RRM1	uco01lyw.4	chr11	+	4130875	C	T	11	16	4130905	D	G	outside
RPL7A	uco04cde.1	chr9	+	136216475	G	A	8	45	136215874	K	R	outside
RPL14	uco03ckg.3	chr3	+	40503437	A	G	8	32	40503686	K	R	outside
RPL14	uco03ckg.3	chr3	+	40503437	A	G	8	32	40503583	S	G	outside
RPL13A	uco02pny.3	chr19	+	49993145	C	T	9	29	49993521	I	V	outside
RNF213	uco02jyh.2	chr17	+	78320687	C	T	9	25	78320200	I	V	outside
RNF213	uco02jyh.2	chr17	+	78320333	G	A	10	18	78320200	I	V	outside
RHOQ	uco02rva.3	chr2	+	46803779	G	A	10	30	46803740	N	S	outside
RBM12B	uco03yfz.3	chr8	-	94746676	C	T	9	14	94745956	K	E	outside
RAP1GDS1	uco03htx.4	chr4	+	99338031	T	C	10	26	99337976	K	E	outside
RAD51AP2	uco10exn.1	chr2	-	17699184	T	C	11	15	17699634	T	A	outside
RAD51AP2	uco10exn.1	chr2	-	17699550	C	A	12	26	17699634	T	A	outside
RABEP1	uco02gbm.4	chr17	+	5264855	A	G	10	32	5264596	M	V	outside
RABEP1	uco02gbm.4	chr17	+	5264645	C	A	8	20	5264596	M	V	outside
PTBP1	uco02lpp.2	chr19	+	804371	C	T	9	29	804407	E	G	outside
PSMD4	uco01exl.3	chr1	+	151239680	C	A	13	36	151239775	T	A	outside
PPP2R5D	uco21yzq.1	chr6	+	42974674	C	T	8	29	42974342	K	E	outside
PPM1B	uco02rtx.3	chr2	+	44445124	C	T	10	16	44445216	K	E	outside
PPAP2A	uco03jpz.3	chr5	-	54721147	G	A	8	41	54721056	Y	C	outside
PPAP2A	uco03jpz.3	chr5	-	54721147	G	A	8	41	54721059	H	R	outside
PPAP2A	uco03jpz.3	chr5	-	54721147	G	A	8	41	54721069	T	A	outside
PPAP2A	uco03jpz.3	chr5	-	54721147	G	A	8	41	54721072	T	A	outside
PPAP2A	uco03jpz.3	chr5	-	54721147	G	A	8	41	54721078	T	A	outside
PPAP2A	uco03jpz.3	chr5	-	54721147	G	A	8	41	54721080	E	G	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
PPAP2A	uc003jpz.3	chr5	-	54721147	G	A	8	41	54721093	T	A	outside
PPAP2A	uc003jpz.3	chr5	-	54721147	G	A	8	41	54721110	K	R	outside
PPAP2A	uc003jpz.3	chr5	-	54721147	G	A	8	41	54721114	R	G	outside
PPAP2A	uc003jpz.3	chr5	-	54721147	G	A	8	41	54721120	K	E	outside
PNMT	uc002hsi.1	chr17	+	37825906	C	T	9	19	37826529	S	G	outside
PLIN4	uc002mar.1	chr19	-	4511676	T	C	11	15	4511783	K	R	outside
PLEKHM2	uc010obo.2	chr1	+	16053733	C	T	11	19	16053951	M	V	outside
PLEKHA8P1	uc001rom.2	chr12	-	45567134	G	T	9	21	45567674	R	G	outside
PLEKHA8P1	uc001rom.2	chr12	-	45567284	C	T	9	23	45567674	R	G	outside
PLEKHA8P1	uc001rom.2	chr12	-	45567134	G	T	9	21	45567409	E	G	outside
PLEKHA8P1	uc001rom.2	chr12	-	45567284	C	T	9	23	45567409	E	G	outside
PKP3	uc021qbk.1	chr11	+	396668	G	A	9	18	397348	S	G	outside
PKP3	uc021qbk.1	chr11	+	397373	G	A	8	17	397348	S	G	outside
PFKM	uc001rrb.2	chr12	+	48528575	C	T	8	52	48528053	H	R	outside
PEMT	uc002grl.3	chr17	-	17412830	G	A	8	14	17412791	M	V	outside
PDE8B	uc003kfa.3	chr5	+	76717683	G	A	8	17	76717776	E	G	outside
PDE8B	uc003kfa.3	chr5	+	76717683	G	A	8	17	76717745	N	D	outside
PDE8B	uc003kfa.3	chr5	+	76717683	G	A	8	17	76717728	K	R	outside
PDE8B	uc003kfa.3	chr5	+	76717683	G	A	8	17	76717727	K	E	outside
PAK1IP1	uc003mzg.3	chr6	+	10697603	C	A	9	23	10697686	K	E	outside
PABPC1	uc003yjs.1	chr8	-	101716543	C	T	10	34	101716611	E	G	outside
PABPC1	uc003yjs.1	chr8	-	101721869	C	T	11	24	101721935	K	E	outside
PABPC1	uc003yjs.1	chr8	-	101727765	A	G	8	28	101727731	D	G	outside
OSGEP	uc001vxf.3	chr14	-	20916905	C	T	9	23	20917172	I	V	outside
OSBPL7	uc002ilx.1	chr17	-	45886547	C	T	8	23	45886233	Q	R	outside
ORAI2	uc010lhz.1	chr7	+	102087009	G	A	9	12	102086979	Q	R	outside
OR51I1	uc010qze.2	chr11	-	5461841	C	A	8	28	5462456	N	D	outside
OR51I1	uc010qze.2	chr11	-	5461883	C	A	9	26	5462456	N	D	outside
OR51I1	uc010qze.2	chr11	-	5462038	C	A	8	19	5462456	N	D	outside
OPTN	uc001ilu.1	chr10	+	13154490	G	T	14	14	13154561	I	V	outside
OIP5	uc001znp.3	chr15	-	41624162	C	G	8	19	41624135	I	V	outside
OGFOD2	uc001udz.1	chr12	+	123463025	G	T	12	27	123463102	M	V	outside
NUPL2	uc003svu.3	chr7	+	23240123	T	G	8	19	23240267	K	R	outside
NUP214	uc004cag.3	chr9	+	134006207	G	A	9	24	134006167	K	E	outside
NUMB	uc001xny.1	chr14	-	73753995	C	T	10	44	73753878	T	A	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
NSUN5P2	uc003twn.3	chr7	-	72419927	C	T	11	19	72419630	R	G	outside
NSUN5P2	uc003twn.3	chr7	-	72419927	C	T	11	19	72419639	M	V	outside
NSUN5P2	uc003twn.3	chr7	-	72419432	C	T	10	18	72419639	M	V	outside
NSUN5P2	uc003twn.3	chr7	-	72419432	C	T	10	18	72419630	R	G	outside
NSUN5P2	uc003twn.3	chr7	-	72419432	C	T	10	18	72419504	M	V	outside
NSUN5P2	uc003twn.3	chr7	-	72419432	C	T	10	18	72419479	N	S	outside
NSUN5P2	uc003twn.3	chr7	-	72419432	C	T	10	18	72419476	E	G	outside
NPM1	uc003mbi.3	chr5	+	170818322	T	G	9	43	170818402	T	A	outside
NOVA1	uc001wpy.3	chr14	-	26918157	A	T	8	13	26917515	T	A	outside
NOVA1	uc001wpy.3	chr14	-	26918157	A	T	8	13	26917530	S	G	outside
NOVA1	uc001wpy.3	chr14	-	26917260	A	G	8	24	26917530	S	G	outside
NOVA1	uc001wpy.3	chr14	-	26917260	A	G	8	24	26917515	T	A	outside
NOVA1	uc001wpy.3	chr14	-	26917692	A	G	12	12	26917515	T	A	outside
NOVA1	uc001wpy.3	chr14	-	26917692	A	G	12	12	26917530	S	G	outside
NME1-NME2	uc002itj.3	chr17	+	49247314	C	T	9	20	49247409	R	G	outside
NFATC1	uc002lnd.3	chr18	+	77221350	G	T	10	53	77221311	D	G	outside
NEIL1	uc002bad.3	chr15	+	75646182	G	A	11	17	75646086	K	R	outside
NBN	uc003yej.1	chr8	-	90993657	C	A	11	39	90993751	S	G	outside
NBEAL2	uc003cqp.3	chr3	+	47049873	G	A	9	25	47049345	Y	C	outside
NBEAL2	uc003cqp.3	chr3	+	47049086	G	A	13	32	47049345	Y	C	outside
NAGLU	uc002hzv.3	chr17	+	40695487	G	A	8	13	40695955	Y	C	outside
MYO9A	uc002atl.4	chr15	-	72119283	C	T	13	22	72119073	K	E	outside
MYO9A	uc002atl.4	chr15	-	72119283	C	T	13	22	72119081	E	G	outside
MYO9A	uc002atl.4	chr15	-	72119271	G	A	13	22	72119073	K	E	outside
MYO9A	uc002atl.4	chr15	-	72119271	G	A	13	22	72119081	E	G	outside
MYO6	uc003pih.1	chr6	+	76596584	C	T	10	15	76596650	E	G	outside
MUC16	uc002mkp.3	chr19	-	8994500	C	T	12	12	8994190	N	S	outside
MSI2	uc010wnm.1	chr17	+	55478804	A	C	9	18	55478740	M	V	outside
MRPS22	uc003etb.3	chr3	+	139075786	G	A	10	30	139075825	E	G	outside
MRPS22	uc003etb.3	chr3	+	139075786	G	A	10	30	139075819	Y	C	outside
MRPS22	uc003etb.3	chr3	+	139075786	G	A	10	30	139075806	T	A	outside
MRPS15	uc001cas.2	chr1	-	36921814	C	T	8	23	36921393	Q	R	outside
MRPS15	uc001cas.2	chr1	-	36921814	C	T	8	23	36921441	N	S	outside
MRPL37	uc001cxa.4	chr1	+	54681839	C	A	9	81	54681937	N	D	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
MRPL37	uc001cxa.4	chr1	+	54681839	C	A	9	81	54681935	K	R	outside
MRPL37	uc001cxa.4	chr1	+	54681839	C	A	9	81	54681934	K	E	outside
MRPL37	uc001cxa.4	chr1	+	54681869	A	G	9	81	54681937	N	D	outside
MRPL37	uc001cxa.4	chr1	+	54681869	A	G	9	81	54681935	K	R	outside
MRPL37	uc001cxa.4	chr1	+	54681869	A	G	9	81	54681934	K	E	outside
MRPL37	uc001cxa.4	chr1	+	54681839	C	A	9	81	54681896	Q	R	outside
MRPL37	uc001cxa.4	chr1	+	54681869	A	G	9	81	54681896	Q	R	outside
MON1B	uc002fez.3	chr16	+	77228730	G	A	13	39	77228232	E	G	outside
MEX3B	uc002bgq.1	chr15	-	82336307	C	T	10	19	82336645	Q	R	outside
MCM8	uc002wmi.3	chr20	+	5966773	A	T	10	29	5966641	Y	C	outside
MAPK8IP3	uc002cmk.3	chr16	+	1818371	G	A	9	23	1818604	E	G	outside
MAP4	uc003csb.2	chr3	-	47957987	G	A	10	41	47958119	K	E	outside
MAP2K2	uc002lzk.3	chr19	-	4102448	C	T	8	36	4102415	K	E	outside
MAP1B	uc003kbw.4	chr5	+	71493792	C	A	8	19	71494320	Q	R	outside
MAP1B	uc003kbw.4	chr5	+	71494449	C	A	8	21	71494320	Q	R	outside
MAP1B	uc003kbw.4	chr5	+	71494308	G	A	8	13	71494320	Q	R	outside
MAGEB10	uc004dbw.3	chrX	+	27839884	C	T	10	17	27839626	Q	R	outside
MAGEB10	uc004dbw.3	chrX	+	27839878	C	A	10	17	27839626	Q	R	outside
LUZP1	uc001bgm.1	chr1	-	23419152	C	A	9	26	23419911	N	D	outside
LUZP1	uc001bgm.1	chr1	-	23420202	A	G	10	21	23419911	N	D	outside
LRRC8D	uc001dnm.3	chr1	+	90400108	G	A	8	14	90399694	H	R	outside
LPCAT4	uc001zig.3	chr15	-	34656451	A	T	11	27	34656406	K	E	outside
LOXL2	uc003xdh.1	chr8	-	23179789	C	T	10	14	23179741	I	V	outside
LILRB4	uc002qgp.3	chr19	+	55179364	G	A	8	23	55179087	Q	R	outside
LILRB4	uc002qgp.3	chr19	+	55179358	C	T	8	23	55179087	Q	R	outside
KRIT1	uc003ulu.1	chr7	-	91842553	A	G	10	20	91842514	T	A	outside
KLHL6	uc003flr.3	chr3	-	183211911	C	A	8	24	183212064	K	E	outside
KIF4B	uc010jih.1	chr5	+	154396190	G	T	9	15	154396267	K	E	outside
KIAA1731	uc009ywb.1	chr11	+	93460148	A	G	9	14	93460523	Q	R	outside
KIAA1731	uc009ywb.1	chr11	+	93460148	A	G	9	14	93460474	S	G	outside
KIAA1429	uc003ygo.2	chr8	-	95504049	T	G	9	29	95504004	R	G	outside
KIAA0913	uc001jve.3	chr10	+	75557018	G	A	9	58	75557200	R	G	outside
KIAA0913	uc001jve.3	chr10	+	75557342	G	A	11	33	75557200	R	G	outside
KIAA0913	uc001jve.3	chr10	+	75557174	G	A	8	13	75557200	R	G	outside
KDM2B	uc001uat.3	chr12	-	121986849	G	A	9	25	121986782	K	R	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
KDM2B	uc001uat.3	chr12	-	121986840	G	A	9	25	121986782	K	R	outside
KCNK9	uc003yvf.1	chr8	-	140630836	C	A	8	33	140631156	E	G	outside
KCNK9	uc003yvf.1	chr8	-	140631331	C	T	8	50	140631156	E	G	outside
KCNK9	uc003yvf.1	chr8	-	140631280	G	T	8	35	140631156	E	G	outside
KCNK9	uc003yvf.1	chr8	-	140631274	C	A	11	32	140631156	E	G	outside
KCNA1	uc001qnh.3	chr12	+	5021548	C	T	8	23	5021742	I	V	outside
KCNA1	uc001qnh.3	chr12	+	5021590	G	A	9	24	5021742	I	V	outside
KCNA1	uc001qnh.3	chr12	+	5021827	C	T	9	23	5021742	I	V	outside
IVNS1ABP	uc001grl.3	chr1	-	185269160	T	C	8	46	185269202	K	R	outside
ITPKB	uc010pvo.2	chr1	-	226924301	G	C	8	38	226924414	Q	R	outside
ITIH3	uc003dfv.2	chr3	+	52837987	C	T	8	41	52837888	E	G	outside
ISG20L2	uc001fpt.1	chr1	-	156697306	A	G	10	35	156696699	Q	R	outside
ISG20L2	uc001fpt.1	chr1	-	156697246	G	A	10	44	156696699	Q	R	outside
IRAK1	uc004fjr.1	chrX	-	153278821	C	A	11	31	153279602	H	R	outside
INHBA	uc003thr.3	chr7	-	41730045	C	T	8	14	41729544	S	G	outside
INHBA	uc003thr.3	chr7	-	41730009	C	A	8	12	41729544	S	G	outside
IL4R	uc002don.3	chr16	+	27374478	T	A	8	30	27375101	K	E	outside
IL4R	uc002don.3	chr16	+	27374556	A	G	12	42	27375101	K	E	outside
IL10RB	uc002yrk.1	chr21	+	34652156	T	C	9	28	34652212	T	A	outside
HSPA5	uc004bpn.3	chr9	-	128001606	G	A	11	21	128000937	N	S	outside
HSPA5	uc004bpn.3	chr9	-	128001606	G	A	11	21	128000938	N	D	outside
HSPA5	uc004bpn.3	chr9	-	128001606	G	A	11	21	128000939	I	M	outside
HSPA5	uc004bpn.3	chr9	-	128001525	C	T	8	21	128000937	N	S	outside
HSPA5	uc004bpn.3	chr9	-	128001525	C	T	8	21	128000938	N	D	outside
HSPA5	uc004bpn.3	chr9	-	128001525	C	T	8	21	128000939	I	M	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778093	S	G	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778129	K	E	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778140	K	R	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778147	R	G	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778167	K	R	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778173	K	R	outside
HSPA1L	uc010jte.3	chr6	-	31778312	G	A	8	59	31778459	T	A	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778224	K	R	outside
HSPA1L	uc010jte.3	chr6	-	31778369	G	A	8	14	31778225	K	E	outside
HSPA1L	uc010jte.3	chr6	-	31778312	G	A	8	59	31778456	T	A	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
HSPA1L	uco10jte.3	chr6	-	31778312	G	A	8	59	31778443	N	S	outside
HSPA1L	uco10jte.3	chr6	-	31778336	G	A	8	15	31778459	T	A	outside
HSPA1L	uco10jte.3	chr6	-	31778336	G	A	8	15	31778456	T	A	outside
HSPA1L	uco10jte.3	chr6	-	31778312	G	A	8	59	31778426	I	V	outside
HSPA1L	uco10jte.3	chr6	-	31778336	G	A	8	15	31778443	N	S	outside
HSPA1L	uco10jte.3	chr6	-	31778362	C	T	8	59	31778459	T	A	outside
HSPA1L	uco10jte.3	chr6	-	31778362	C	T	8	59	31778456	T	A	outside
HSPA1L	uco10jte.3	chr6	-	31778336	G	A	8	15	31778426	I	V	outside
HSPA1L	uco10jte.3	chr6	-	31778362	C	T	8	59	31778443	N	S	outside
HSPA1L	uco10jte.3	chr6	-	31778369	G	A	8	14	31778443	N	S	outside
HSPA1L	uco10jte.3	chr6	-	31778362	C	T	8	59	31778426	I	V	outside
HSPA1L	uco10jte.3	chr6	-	31778369	G	A	8	14	31778426	I	V	outside
HSPA1L	uco10jte.3	chr6	-	31778435	C	A	8	15	31778459	T	A	outside
HSPA1L	uco10jte.3	chr6	-	31778435	C	A	8	15	31778456	T	A	outside
HSPA1L	uco10jte.3	chr6	-	31778435	C	A	8	15	31778443	N	S	outside
HSP90B1	uco01tkb.1	chr12	+	104340404	G	A	9	42	104340469	I	V	outside
HSP90B1	uco01tkb.1	chr12	+	104340404	G	A	9	42	104340451	R	G	outside
HSP90AB1	uco03oxa.1	chr6	+	44220000	G	A	10	40	44219939	S	G	outside
HSP90AB1	uco03oxa.1	chr6	+	44217142	G	C	8	22	44217171	K	E	outside
HSP90AB1	uco03oxa.1	chr6	+	44217154	G	A	8	22	44217171	K	E	outside
HSP90AA1	uco01ykv.4	chr14	-	102550183	G	T	8	14	102550216	K	E	outside
HNRNPA3	uco02ulc.1	chr2	+	178081255	G	A	11	22	178081272	K	E	outside
HNRNPA3	uco02ulc.1	chr2	+	178081264	T	C	11	22	178081272	K	E	outside
HMGXB3	uco03lrk.4	chr5	+	149403937	T	C	9	13	149403901	Q	R	outside
HLA-G	uco03nnw.2	chr6	+	29795680	C	A	10	14	29796093	S	G	outside
HLA-G	uco03nnw.2	chr6	+	29796590	C	T	11	60	29796554	H	R	outside
HIST2H2AC	uco01etd.3	chr1	+	149858592	C	T	8	18	149858879	K	E	outside
HIST2H2AB	uco01ete.3	chr1	-	149859427	T	C	13	27	149859202	R	G	outside
HIST2H2AB	uco01ete.3	chr1	-	149859427	T	C	13	27	149859229	I	V	outside
HIST2H2AB	uco01ete.3	chr1	-	149859427	T	C	13	27	149859238	T	A	outside
GTF3C1	uco02dov.2	chr16	-	27472721	G	T	11	23	27472763	M	V	outside
GTF3C1	uco02dov.2	chr16	-	27472724	G	A	11	23	27472763	M	V	outside
GSK3B	uco03edn.3	chr3	-	119562179	C	T	9	12	119562119	N	S	outside
GSK3B	uco03edn.3	chr3	-	119562179	C	T	9	12	119562120	N	D	outside
GRIA3	uco04etq.4	chrX	+	122598732	G	T	8	13	122598962	R	G	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
GPRIN2	uco01jec.3	chr10	+	46999542	G	T	10	54	46999805	R	G	outside
GPRIN2	uco01jec.3	chr10	+	47000010	C	T	10	23	46999805	R	G	outside
GPRIN2	uco01jec.3	chr10	+	46999935	G	A	8	22	46999805	R	G	outside
GPRIN2	uco01jec.3	chr10	+	46999689	G	A	10	62	46999805	R	G	outside
GPRIN2	uco01jec.3	chr10	+	46999755	C	T	10	62	46999805	R	G	outside
GPR160	uco03fgi.3	chr3	+	169802047	A	T	12	27	169801864	K	R	outside
GP2	uco02dgv.3	chr16	-	20335150	G	T	8	30	20335369	R	G	outside
GP2	uco02dgv.3	chr16	-	20335420	G	A	10	23	20335369	R	G	outside
GP2	uco02dgv.3	chr16	-	20331721	G	A	9	18	20331702	E	G	outside
GNPTG	uco02clm.3	chr16	+	1412631	G	A	8	34	1412723	T	A	outside
GLUD2	uco04eto.3	chrX	+	120182653	G	A	9	14	120183106	H	R	outside
GLUD2	uco04eto.3	chrX	+	120181768	C	T	8	12	120182218	N	S	outside
GLUD2	uco04eto.3	chrX	+	120181768	C	T	8	12	120181813	Q	R	outside
GLRX	uco21ybn.1	chr5	-	95152246	G	T	8	23	95152221	Q	R	outside
GIPC1	uco02myt.3	chr19	-	14593638	T	C	9	19	14593605	T	A	outside
GFOD2	uco02eub.3	chr16	-	67709708	C	T	11	15	67709489	M	V	outside
GFOD2	uco02eub.3	chr16	-	67709585	C	G	9	15	67709489	M	V	outside
GALNTL4	uco01mjo.2	chr11	-	11292699	C	G	11	47	11292815	K	R	outside
G6PC3	uco02iey.3	chr17	+	42152339	C	A	9	21	42153086	E	G	outside
FPR2	uco02pxr.3	chr19	+	52272375	C	A	9	30	52272914	T	A	outside
FPR2	uco02pxr.3	chr19	+	52272375	C	A	9	30	52272912	D	G	outside
FPR2	uco02pxr.3	chr19	+	52272375	C	A	9	30	52272909	N	S	outside
FPR2	uco02pxr.3	chr19	+	52272396	C	G	9	30	52272914	T	A	outside
FPR2	uco02pxr.3	chr19	+	52272396	C	G	9	30	52272912	D	G	outside
FPR2	uco02pxr.3	chr19	+	52272396	C	G	9	30	52272909	N	S	outside
FPR2	uco02pxr.3	chr19	+	52272375	C	A	9	30	52272869	S	G	outside
FPR2	uco02pxr.3	chr19	+	52272375	C	A	9	30	52272866	T	A	outside
FPR2	uco02pxr.3	chr19	+	52272396	C	G	9	30	52272869	S	G	outside
FPR2	uco02pxr.3	chr19	+	52272396	C	G	9	30	52272866	T	A	outside
FPR2	uco02pxr.3	chr19	+	52272594	C	T	10	37	52272869	S	G	outside
FPR2	uco02pxr.3	chr19	+	52272594	C	T	10	37	52272866	T	A	outside
FOXN2	uco02rwh.1	chr2	+	48602252	C	T	8	23	48602197	Y	C	outside
FLOT1	uco03nrm.3	chr6	-	30708048	C	T	11	12	30708311	I	V	outside
FGFBP2	uco03gon.3	chr4	-	15964088	C	A	9	29	15964334	N	S	outside
FGFBP2	uco03gon.3	chr4	-	15964356	C	T	10	14	15964334	N	S	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
FDPS	uc001fkc.2	chr1	+	155279984	C	A	10	27	155279696	Y	C	outside
FDPS	uc001fkc.2	chr1	+	155279944	G	A	8	15	155279696	Y	C	outside
FASN	uc002kdu.3	chr17	-	80041513	C	T	11	28	80041164	K	E	outside
FAM82A2	uc001zmq.1	chr15	-	41029864	C	T	8	18	41030198	M	V	outside
FAM208A	uc003die.4	chr3	-	56680634	G	A	8	62	56680561	H	R	outside
FAM149A	uc003iyt.4	chr4	+	187077231	G	T	12	19	187077238	D	E	outside
EXOC8	uc001huq.3	chr1	-	231471715	G	A	13	24	231471358	R	G	outside
EXOC4	uc003vrk.3	chr7	+	132959841	G	A	8	14	132959897	I	V	outside
ETAA1	uc002sdz.1	chr2	+	67630779	G	A	8	18	67630739	S	G	outside
ERP44	uc004bam.3	chr9	-	102768840	C	T	10	19	102768887	H	R	outside
ERGIC2	uc001riv.3	chr12	-	29510634	A	C	9	14	29510619	H	R	outside
EPB41L2	uc003qch.2	chr6	-	131216112	G	A	9	33	131216166	K	E	outside
EMILIN2	uc002kln.3	chr18	+	2892353	G	A	10	18	2891758	K	E	outside
EMILIN2	uc002kln.3	chr18	+	2891303	G	A	8	21	2891758	K	E	outside
EIF4G3	uc001bef.3	chr1	-	21268371	G	A	8	41	21268040	Q	R	outside
EIF4G3	uc001bef.3	chr1	-	21212828	C	G	10	20	21212695	Q	R	outside
EIF4G3	uc001bef.3	chr1	-	21268167	T	C	8	41	21268040	Q	R	outside
EIF4ENIF1	uc003ala.2	chr22	-	31850207	C	T	9	13	31850299	D	G	outside
EIF4E	uc011ceb.1	chr4	-	99823085	G	T	8	41	99823132	E	G	outside
EGFR	uc003tqk.3	chr7	+	55221768	C	T	9	20	55221794	N	D	outside
EDC3	uc002ayn.3	chr15	-	74964111	A	T	9	32	74963882	Q	R	outside
ECE2	uc003fni.4	chr3	+	183975498	C	T	10	22	183975372	H	R	outside
E2F1	uc002wzu.4	chr20	-	32265234	G	A	10	14	32264785	E	G	outside
DYNC1H1	uc001yks.2	chr14	+	102515891	T	C	9	29	102516132	S	G	outside
DYNC1H1	uc001yks.2	chr14	+	102516208	C	T	9	27	102516132	S	G	outside
DUSP7	uc003dct.3	chr3	-	52088013	C	A	8	16	52087995	N	D	outside
DUOXA1	uc010bec.3	chr15	-	45412433	G	C	9	16	45412889	D	G	outside
DUOXA1	uc010bec.3	chr15	-	45412433	G	C	9	16	45412865	K	R	outside
DUOXA1	uc010bec.3	chr15	-	45412433	G	C	9	16	45412854	R	G	outside
DSTN	uc002wpq.3	chr20	+	17581486	C	T	11	15	17581435	K	R	outside
DHX38	uc002fcb.3	chr16	+	72130067	C	T	10	30	72130372	K	E	outside
DHX29	uc003jpx.3	chr5	-	54565396	C	T	8	14	54565393	N	D	outside
DCN	uc001tbu.3	chr12	-	91572134	C	T	12	41	91572222	I	M	outside
DCN	uc001tbu.3	chr12	-	91572134	C	T	12	41	91572205	D	G	outside
DCAF8L2	uc011mjy.2	chrX	+	27765977	C	T	10	55	27766223	N	S	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
DCAF8L2	uco11mjy.2	chrX	+	27765977	C	T	10	55	27766217	K	R	outside
DCAF8L2	uco11mjy.2	chrX	+	27765977	C	T	10	55	27766213	K	E	outside
DCAF8L2	uco11mjy.2	chrX	+	27765977	C	T	10	55	27766189	I	V	outside
DCAF8L2	uco11mjy.2	chrX	+	27765977	C	T	10	55	27766186	R	G	outside
DCAF8L2	uco11mjy.2	chrX	+	27765977	C	T	10	55	27766172	Q	R	outside
DCAF16	uco03gpn.3	chr4	-	17805218	G	A	8	14	17805279	I	M	outside
DBNL	uco03tjo.4	chr7	+	44099663	T	C	12	16	44099042	R	G	outside
DAPK1	uco04apc.3	chr9	+	90296515	T	A	8	21	90296538	T	A	outside
CXorf65	uco11mpo.2	chrX	-	70323890	C	T	8	38	70324174	I	V	outside
CUL9	uco03ouk.3	chr6	+	43181258	G	T	8	14	43180970	K	E	outside
CTNNA3	uco01jmw.2	chr10	-	68979439	A	C	8	15	68979490	T	A	outside
CTNNA3	uco01jmw.2	chr10	-	68979520	T	C	11	20	68979490	T	A	outside
CSNK1D	uco02kei.3	chr17	-	80211075	G	T	8	14	80210903	H	R	outside
CSNK1A1	uco03lqw.1	chr5	-	148929645	C	A	9	20	148929719	Q	R	outside
CSNK1A1	uco03lqw.1	chr5	-	148929645	C	A	9	20	148929675	K	E	outside
CSNK1A1	uco03lqw.1	chr5	-	148929645	C	A	9	20	148929674	K	R	outside
CSDE1	uco01efi.3	chr1	-	115262216	T	C	8	29	115261356	K	E	outside
CREB1	uco02vcc.3	chr2	+	208440158	C	T	8	31	208440032	N	S	outside
CREB1	uco02vcc.3	chr2	+	208440098	G	T	8	28	208440032	N	S	outside
CREB1	uco02vcc.3	chr2	+	208440098	G	T	8	28	208440076	T	A	outside
CRB2	uco04bnx.1	chr9	+	126133110	G	A	8	18	126132620	T	A	outside
CPPED1	uco02dca.4	chr16	-	12798544	C	T	8	19	12798508	K	E	outside
CLTC	uco02ixq.1	chr17	+	57721821	T	C	10	35	57721808	I	V	outside
CLPX	uco02aom.3	chr15	-	65458971	G	A	8	14	65459090	H	R	outside
CIT	uco01txj.2	chr12	-	120150420	C	A	9	43	120150152	D	G	outside
CIT	uco01txj.2	chr12	-	120150411	A	G	9	46	120150152	D	G	outside
CIT	uco01txj.2	chr12	-	120150099	C	A	8	26	120150152	D	G	outside
CIT	uco01txj.2	chr12	-	120150105	C	T	8	26	120150152	D	G	outside
CFH	uco01gtj.4	chr1	+	196706041	A	G	8	50	196706769	S	G	outside
CFH	uco01gtj.4	chr1	+	196706689	G	C	8	21	196706769	S	G	outside
CES2	uco02eqr.3	chr16	+	66976557	G	T	9	32	66976682	N	D	outside
CEP250	uco21wco.1	chr20	+	34092253	G	A	11	15	34092511	Q	R	outside
CEP192	uco10xac.2	chr18	+	13073109	C	A	8	19	13073044	S	G	outside
CEP192	uco10xac.2	chr18	+	13073084	C	T	9	23	13073044	S	G	outside
CENPI	uco11mrg.2	chrX	+	100381783	G	A	9	23	100382236	N	S	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
CDK13	uco03thh.4	chr7	+	40132510	G	A	8	18	40132494	I	V	outside
CD226	uco02lkm.4	chr18	-	67563246	C	T	12	25	67563158	Q	R	outside
CCDC40	uco10dht.3	chr17	+	78061472	C	T	10	21	78061883	N	S	outside
CCDC40	uco10dht.3	chr17	+	78023867	G	T	10	16	78024008	K	R	outside
CCDC159	uco10xlt.2	chr19	+	11462590	C	T	10	27	11462733	Q	R	outside
CASP8	uco02uxr.1	chr2	+	202131409	C	T	8	35	202131304	Q	R	outside
CADPS	uco03dll.2	chr3	-	62423880	C	A	8	20	62423807	E	G	outside
CADPS	uco03dll.2	chr3	-	62423877	G	A	8	20	62423807	E	G	outside
CADPS	uco03dll.2	chr3	-	62423865	C	A	8	20	62423807	E	G	outside
C5orf41	uco03mch.3	chr5	+	172517394	G	T	11	15	172517877	K	R	outside
C4orf29	uco21xrt.1	chr4	+	128949941	C	A	10	17	128949892	N	S	outside
C3AR1	uco01qtv.1	chr12	-	8212520	G	A	8	14	8212760	T	A	outside
C16orf88	uco02dgq.3	chr16	-	19718306	C	T	10	44	19718239	E	G	outside
BTAF1	uco01khr.3	chr10	+	93768665	T	G	9	14	93767971	K	R	outside
BRD9	uco03jbq.3	chr5	-	891952	C	T	8	27	891832	K	E	outside
BRD9	uco03jbq.3	chr5	-	891952	C	T	8	27	891834	H	R	outside
BOD1L	uco03gmz.1	chr4	-	13601770	C	T	9	42	13601238	E	G	outside
BOD1L	uco03gmz.1	chr4	-	13603498	A	G	13	17	13603464	D	G	outside
ATHL1	uco10qvu.2	chr11	+	293449	A	G	10	25	292983	K	R	outside
ATHL1	uco10qvu.2	chr11	+	292634	G	T	10	25	292983	K	R	outside
ATHL1	uco10qvu.2	chr11	+	292926	C	T	9	26	292983	K	R	outside
ATG2A	uco010bx.3	chr11	-	64663978	C	G	10	43	64663957	N	D	outside
ATAD3B	uco01afv.3	chr1	+	1425948	G	A	11	43	1425783	K	R	outside
ARID4B	uco01hwq.3	chr1	-	235377133	C	T	8	32	235377328	K	E	outside
ARID4B	uco01hwq.3	chr1	-	235377154	T	C	9	21	235377328	K	E	outside
ARID4B	uco01hwq.3	chr1	-	235377184	A	G	8	16	235377328	K	E	outside
ARHGEF3	uco03dih.2	chr3	-	56763414	A	G	10	28	56763354	M	V	outside
ARHGAP33	uco020bt.2	chr19	+	36278502	G	A	14	23	36278727	Y	C	outside
APOL3	uco03aou.3	chr22	-	36537325	G	A	9	22	36537631	N	D	outside
APOL3	uco03aou.3	chr22	-	36537859	C	A	11	18	36537631	N	D	outside
ANKHD1	uco03lfr.3	chr5	+	139876771	T	C	9	32	139876842	T	A	outside
ALPP	uco02vsq.3	chr2	+	233246012	C	T	8	21	233245190	T	A	outside
ALPP	uco02vsq.3	chr2	+	233246012	C	T	8	21	233245729	S	G	outside
ALPK1	uco03ian.4	chr4	+	113353162	C	A	12	22	113352432	S	G	outside
ALPK1	uco03ian.4	chr4	+	113353162	C	A	12	22	113352666	N	D	outside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
ALDOC	uc002hbp.3	chr17	-	26902292	C	T	9	44	26902513	K	R	outside
ALDH1L1	uc003eim.1	chr3	-	125824635	G	T	8	17	125824596	K	E	outside
AKAP2	uc011lwi.2	chr9	+	112900355	G	A	10	53	112900457	E	G	outside
AIM2	uc001ftj.1	chr1	-	159035951	C	T	13	13	159035969	T	A	outside
ADAMTS12	uc003jia.1	chr5	-	33751524	A	C	9	18	33751586	E	G	outside
ADAMTS12	uc003jia.1	chr5	-	33751524	A	C	9	18	33751577	Y	C	outside
ADAMTS12	uc003jia.1	chr5	-	33751524	A	C	9	18	33751574	H	R	outside
ADAMTS12	uc003jia.1	chr5	-	33751590	G	T	8	28	33751568	H	R	outside
ADAMTS12	uc003jia.1	chr5	-	33751569	G	A	9	18	33751586	E	G	outside
ADAMTS12	uc003jia.1	chr5	-	33751569	G	A	9	18	33751577	Y	C	outside
ADAMTS12	uc003jia.1	chr5	-	33751569	G	A	9	18	33751574	H	R	outside
ADAM20	uc001xme.3	chr14	-	70990832	C	T	8	28	70990511	N	D	outside
ADAM20	uc001xme.3	chr14	-	70990283	G	A	9	19	70990511	N	D	outside
ADAM20	uc001xme.3	chr14	-	70990448	C	T	10	38	70990511	N	D	outside
ACAP3	uc001aeb.2	chr1	-	1237393	G	T	8	58	1237368	E	G	outside
ABCA13	uc003toq.2	chr7	+	48314626	G	A	8	22	48315009	T	A	outside
AAK1	uc002sfp.2	chr2	-	69741764	G	A	9	23	69741674	M	V	outside
ZNF91	uc002nre.3	chr19	-	23544208	C	T	8	23	23544208	K	E	inside
ZNF83	uc010epz.3	chr19	-	53116803	C	T	8	15	53116839	S	G	inside
ZNF793	uc010efm.3	chr19	+	38028510	G	A	10	41	38028483	E	G	inside
ZNF721	uc003gag.3	chr4	-	435804	G	A	8	43	435906	K	E	inside
ZNF283	uc0020xr.4	chr19	+	44352024	T	C	12	55	44351672	T	A	inside
ZNF283	uc0020xr.4	chr19	+	44352024	T	C	12	55	44351682	Q	R	inside
ZNF283	uc0020xr.4	chr19	+	44352024	T	C	12	55	44351693	T	A	inside
ZNF283	uc0020xr.4	chr19	+	44352024	T	C	12	55	44351711	K	E	inside
ZNF283	uc0020xr.4	chr19	+	44352024	T	C	12	55	44351717	K	E	inside
ZNF283	uc0020xr.4	chr19	+	44352024	T	C	12	55	44351718	K	R	inside
ZCCHC24	uc001kak.3	chr10	-	81192472	C	T	11	20	81192477	Y	C	inside
XPO1	uc002sbj.3	chr2	-	61719601	A	G	9	25	61719583	K	E	inside
TUBG1	uc002ian.3	chr17	+	40765675	C	T	11	13	40765675	D	G	inside
ST3GAL6	uc003dsy.3	chr3	+	98492809	C	A	8	23	98492827	N	S	inside
SPEN	uc001axk.1	chr1	+	16257948	C	G	13	15	16258632	N	S	inside
SPEN	uc001axk.1	chr1	+	16257957	G	A	13	15	16258632	N	S	inside
SNX11	uc010wlg.1	chr17	+	46198815	A	T	10	31	46198763	R	G	inside
SMARCA1	uc004eun.4	chrX	-	128645905	C	A	12	46	128645951	K	E	inside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
SMARCA1	uco04eun.4	chrX	-	128645905	C	A	12	46	128645941	Q	R	inside
SMARCA1	uco04eun.4	chrX	-	128645905	C	A	12	46	128645939	T	A	inside
SMARCA1	uco04eun.4	chrX	-	128645905	C	A	12	46	128645936	I	V	inside
SMARCA1	uco04eun.4	chrX	-	128645951	G	T	12	46	128645936	I	V	inside
SMARCA1	uco04eun.4	chrX	-	128645951	G	T	12	46	128645939	T	A	inside
SMARCA1	uco04eun.4	chrX	-	128645951	G	T	12	46	128645941	Q	R	inside
SMARCA1	uco04eun.4	chrX	-	128649708	C	A	9	14	128649720	N	D	inside
SMARCA1	uco04eun.4	chrX	-	128649711	G	A	9	14	128649720	N	D	inside
SMARCA1	uco04eun.4	chrX	-	128645951	G	T	12	46	128645951	K	E	inside
SLC30A5	uco03jvh.3	chr5	+	68412386	G	A	9	57	68411930	T	A	inside
SLC30A5	uco03jvh.3	chr5	+	68412341	G	A	9	57	68411930	T	A	inside
SF3B2	uco010gy.1	chr11	+	65827044	C	T	12	41	65827019	K	E	inside
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603095	T	A	inside
SEPT4	uco10wny.2	chr17	-	56603080	G	T	9	24	56603085	Y	C	inside
RRNAD1	uco01fpu.3	chr1	+	156703952	G	T	8	34	156703904	Q	R	inside
RPS3A	uco03ilz.3	chr4	+	152020863	A	G	8	16	152020856	K	E	inside
RPL14	uco03ckg.3	chr3	+	40503437	A	G	8	32	40503441	I	M	inside
RALY	uco02xab.3	chr20	+	32666318	C	T	8	23	32666321	Q	R	inside
RALGAPB	uco02xiw.3	chr20	+	37161461	C	T	9	24	37161445	T	A	inside
PHF3	uco03pen.2	chr6	+	64422163	C	T	10	23	64422046	K	R	inside
PABPC1	uco03yjs.1	chr8	-	101724673	T	C	8	46	101724649	I	V	inside
PABPC1	uco03yjs.1	chr8	-	101724630	C	A	8	46	101724649	I	V	inside
NOC3L	uco01kjq.1	chr10	-	96094399	C	A	12	17	96094399	K	E	inside
NCAPG	uco03gpp.3	chr4	+	17842266	T	G	8	41	17842230	E	G	inside
MYO6	uco03pih.1	chr6	+	76596647	C	T	8	13	76596650	E	G	inside
MYNN	uco03ffu.3	chr3	+	169501280	A	T	13	24	169501271	K	R	inside
MRPS22	uco03etb.3	chr3	+	139075786	G	A	10	30	139075770	K	E	inside
MRPS22	uco03etb.3	chr3	+	139075786	G	A	10	30	139075771	K	R	inside
MRPS22	uco03etb.3	chr3	+	139075786	G	A	10	30	139075773	T	A	inside
MRPS22	uco03etb.3	chr3	+	139075786	G	A	10	30	139075795	Y	C	inside
MRPS22	uco03etb.3	chr3	+	139075786	G	A	10	30	139075783	Q	R	inside
MLF1	uco03fbx.3	chr3	+	158317861	T	A	8	21	158317839	S	G	inside
MFSD5	uco01sch.2	chr12	+	53647587	C	G	9	55	53647776	H	R	inside
ME1	uco03pjy.3	chr6	-	84055987	C	T	9	13	84055987	I	V	inside
HSPA1L	uco10jte.3	chr6	-	31778369	G	A	8	14	31778209	K	R	inside

Continued on next page

Table A.4 – continued

Gene name	ID UCSC	Chr	Strand	Silent mut pos	Base	Mut	Longest stem	N° base pairs	Editing position	Gen aa	Edit aa	Where editing
HSPA1L	uco10jte.3	chr6	-	31778369	G	A	8	14	31778210	K	E	inside
HSPA1L	uco10jte.3	chr6	-	31778435	C	A	8	15	31778426	I	V	inside
GSTM5	uco01dyn.3	chr1	+	110256316	C	T	50	87	110256304	K	R	inside
GSTM5	uco01dyn.3	chr1	+	110256316	C	T	50	87	110256306	I	V	inside
FKBP15	uco04bgs.2	chr9	-	115950084	C	G	8	30	115950083	Q	R	inside
FAM111A	uco10rkp.2	chr11	+	58919839	C	A	8	16	58919829	R	G	inside
FAH	uco02bfm.2	chr15	+	80465389	C	T	8	29	80465406	K	E	inside
DYNC1H1	uco01yks.2	chr14	+	102461588	T	C	8	68	102461540	H	R	inside
DYNC1H1	uco01yks.2	chr14	+	102516142	G	A	10	22	102516132	S	G	inside
DGKA	uco01sij.3	chr12	+	56346872	G	A	11	23	56346932	E	G	inside
CTNNAL1	uco04bdo.1	chr9	-	111732693	G	T	8	20	111732686	Q	R	inside
CSE1L	uco02xty.3	chr20	+	47711382	C	T	10	32	47711355	E	G	inside
CRELD2	uco10hal.2	chr22	+	50313437	T	C	10	24	50313458	E	G	inside
CREB1	uco02vcc.3	chr2	+	208440158	C	T	8	31	208440076	T	A	inside
CDK13	uco03thh.4	chr7	+	40127766	G	A	9	55	40127792	M	V	inside
CDK13	uco03thh.4	chr7	+	40127781	G	A	9	39	40127792	M	V	inside
CCNK	uco01ygi.4	chr14	+	99961962	A	G	8	28	99961938	Q	R	inside
CCNDBP1	uco01zqv.3	chr15	+	43483789	G	A	10	48	43483813	K	R	inside
CACNG8	uco02qcs.2	chr19	+	54485586	G	T	13	23	54485579	S	G	inside
ASB1	uco02vyg.3	chr2	+	239355106	G	A	11	36	239355126	I	V	inside
ARRB1	uco010owe.2	chr11	-	74988429	G	T	8	22	74988426	K	E	inside
ARID4B	uco01hwq.3	chr1	-	235377280	G	A	13	61	235377328	K	E	inside
ARID4B	uco01hwq.3	chr1	-	235377292	G	A	13	61	235377328	K	E	inside
ARHGEF3	uco03dih.2	chr3	-	56763414	A	G	10	28	56763440	E	G	inside
ALPP	uco02vsq.3	chr2	+	233246012	C	T	8	21	233245167	Q	R	inside
ALKBH2	uco01tnx.2	chr12	-	109526247	C	A	8	15	109526247	I	V	inside
ADAMTS12	uco03jia.1	chr5	-	33751524	A	C	9	18	33751568	H	R	inside
ADAMTS12	uco03jia.1	chr5	-	33751590	G	T	8	28	33751574	H	R	inside
ADAMTS12	uco03jia.1	chr5	-	33751590	G	T	8	28	33751577	Y	C	inside
ADAMTS12	uco03jia.1	chr5	-	33751590	G	T	8	28	33751586	E	G	inside
ADAMTS12	uco03jia.1	chr5	-	33751569	G	A	9	18	33751568	H	R	inside